

Automatische Strukturierung von Lerneinheiten mit semantischen Methoden am Beispiel der Werkstoffe im Bauwesen

Vom Fachbereich Bauingenieurwesen und Geodäsie der Technischen Universität Darmstadt zur
Erlangung des akademischen Grades eines Doktor-Ingenieurs (Dr.-Ing.) genehmigte Dissertation

von
Dipl.-Ing. Nils Schnittker
aus Bremen

Referent: Prof. Dr.-Ing. Peter Grübl

Korreferent: Prof. Dr.-Ing. Jörg Lange

Tag der Einreichung: 27.06.2008

Tag der mündlichen Prüfung: 12.02.2009

Darmstadt 2009

D17

Vorwort

Diese Arbeit entstand während und nach meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Institut für Massivbau der TU Darmstadt. Sie basiert auf dem unter der Federführung von Herrn Prof. Grübl entwickelten Lernnetzwerk „Werkstoffe im Bauwesen“ (WiBA-Net).

Für die Betreuung der vorliegenden Arbeit danke ich Herrn Prof. Dr.-Ing. Peter Grübl, der mit seinem unermüdlichen Einsatz für das Projekt WiBA-Net und der Förderung des E-Learnings in der Lehre seines Fachgebiets eine wichtige Inspirationsquelle für die vorliegende Arbeit darstellte.

Bei Herrn Prof. Dr.-Ing. Jörg Lange möchte ich mich herzlich für die Übernahme des Korreferats bedanken.

Für die gute Zusammenarbeit möchte ich mich auch bei meinen ehemaligen Kolleginnen und Kollegen am Institut für Massivbau bedanken, insbesondere bei Herrn Dipl.-Ing. Bernd Schmidt.

Danken möchte ich auch meinen Eltern Hans und Ursula Schnittker für ihre Förderung und Unterstützung sowie meiner Schwester Inga, die in der deutschen Rechtschreibung sehr bewandert ist.

Ganz besonderen Dank schulde ich meiner Freundin Mona für ihre Unterstützung und nicht zuletzt für ihre Geduld. Vielen Dank!

Darmstadt, im Juni 2008

Nils Schnittker

We learned more from a three minute record than we ever learned in school.

Bruce Springsteen, „No Surrender“

Kurzfassung

In dieser Arbeit wird ein Programmsystem vorgeschlagen und als Prototyp implementiert, mit dem sowohl Lernende als auch Lehrende automatisch Lerneinheiten generieren können und das zur Wissensvermittlung im Bauingenieurwesen eingesetzt werden kann. Ist der Nutzer ein Hochschullehrer, so kann die erstellte Lerneinheit nach Wunsch überarbeitet und danach den Studierenden zur Verfügung gestellt werden.

Inhaltlich und didaktisch fußt das System auf dem im Rahmen eines Forschungsprojekts entwickelten Lernnetzwerk „Werkstoffe im Bauwesen“ (WiBA-Net). Ein wesentlicher Bestandteil dieses Lernnetzwerks sind die so genannten „Lehrpfade“, strukturierte Lerneinheiten zu einem jeweils abgegrenzten Themenbereich. Sie sind aus einzelnen Wissenseinheiten aufgebaut, ihre Struktur wird von Autoren manuell festgelegt. Das entwickelte System erstellt Lehrpfade analog zum WiBA-Net-Konzept automatisch aus vorhandenen Wissenseinheiten. Dabei wird das didaktische Konzept des WiBA-Net durch den Aspekt des freien Lernens ergänzt.

Das Konzept und die Funktionalität des Systems werden an den Inhalten des WiBA-Net demonstriert, die das Wissensgebiet „Werkstoffe im Bauwesen“ umfassen. Die Inhalte liegen in textlicher und multimedialer Form vor, für die vorliegende Arbeit werden jedoch nur die textlichen Lernelemente benötigt. Aus ihnen werden linguistische und textstatistische Informationen extrahiert. Die Texte und die aus ihnen extrahierten Informationen stellen zusammen die Inhaltsbasis des entwickelten Systems dar. Eine Erweiterung des Systems um zusätzliche Wissensgebiete ist möglich.

Die genutzten linguistischen Informationen umfassen die Rückführung von abgeleiteten Wortformen auf ihre Stämme sowie die Zerlegung von Komposita. Die textstatistischen Informationen beinhalten die Signifikanzen des gemeinsamen Auftretens von Begriffspaaren („Kookkurenzen“) und die Häufigkeiten der einzelnen Begriffe. Diese Informationen ermöglichen die Bestimmung von Ähnlichkeitsmaßen zwischen Dokumenten bzw. Wissenseinheiten sowie die Durchführung einer semantischen Suche.

Bei einer Suchanfrage werden zunächst alle Wissenseinheiten, die das Thema der Anfrage behandeln, identifiziert. Für die gefundenen Dokumente werden mittels impliziter bzw. unscharfer Inferenzmechanismen Relevanzwerte bestimmt, die ihre Zugehörigkeit zum zu erstellenden Lehrpfad charakterisieren. Wissenseinheiten mit hinreichend hoher Relevanz werden dann mit einem Clustering-Verfahren gruppiert und anhand didaktischer Kriterien sortiert.

Abschließend wird aus der Analyse automatisch generierter Lehrpfade eine Methode erarbeitet, aus einem Kennwert, der auf textstatistischen Eigenschaften der Suchanfrage beruht, eine Prognose über die voraussichtliche inhaltliche Qualität des erstellten Lehrpfads abzuleiten.

Der dem Lerner angezeigte Lehrpfad versetzt diesen in die Lage, die durch die vorgegebenen Suchbegriffe implizierte Frage selbst zu beantworten und zukünftig eigene Entscheidungen auf der Basis des erworbenen Wissens zu treffen.

Inhaltsverzeichnis

1 Motivation und Zielsetzung.....	1
1.1 Ausgangssituation.....	1
1.2 Verbesserungspotential beim Einsatz von E-Learning.....	3
1.3 Ziele und Vorgehensweise.....	6
1.4 Aufbau der Arbeit	8
2 Stand des Wissens.....	11
2.1 E-Learning.....	11
2.1.1 Was ist E-Learning?.....	11
2.1.2 Pädagogisch-didaktische Aspekte und Anwendungsbeispiele aus dem E-Learning.....	11
2.1.3 Entwicklung des E-Learning.....	15
2.1.3.1 Historische Entwicklung.....	15
2.1.3.2 Einfluss der Neuen Medien.....	17
2.1.3.3 Standardisierung.....	21
2.1.3.4 Aktuelle Forschungsgebiete.....	22
2.1.4 Spezielle Formen des E-Learnings.....	22
2.1.5 Einsatz von E-Learning an Hochschulen.....	23
2.1.5.1 Allgemeiner Einsatz an Hochschulen.....	23
2.1.5.2 E-Learning Projekte im Bauingenieurwesen.....	25
2.1.6 Zusammenfassung: Aspekte des Einsatzes von E-Learning.....	26
2.2 Information-Retrieval.....	27
2.2.1 Grundlagen des Information-Retrieval.....	27
2.2.1.1 Aufgaben des Information-Retrieval.....	27
2.2.1.2 Relevanz von Suchergebnissen	29
2.2.1.3 Precision und Recall.....	29
2.2.2 Verfahren des Information-Retrieval.....	31
2.2.2.1 Indizierung und Boolesches Retrieval.....	31
2.2.2.2 Vektorraummodell.....	32
2.2.2.3 Erweiterung von Suchanfragen.....	34
2.2.2.4 Linguistische Verfahren.....	35

2.2.2.5 Weitere Verfahren.....	35
2.2.3 Mathematische Grundlagen für das Information-Retrieval.....	36
2.2.3.1 Ähnlichkeitsmaße.....	36
2.2.3.2 Clustering.....	37
2.2.3.3 Fuzzy Logic.....	38
2.3 Sprachwissenschaft.....	42
2.3.1 Linguistische Grundlagen.....	42
2.3.2 Computerlinguistik.....	45
3 Lernnetzwerk „Werkstoffe im Bauwesen“ (WiBA-Net).....	53
3.1 Vorbemerkung.....	53
3.2 Konzept des WiBA-Net.....	53
3.3 Funktionalitäten des WiBA-Net.....	54
3.4 Selbststudium mit Lehrpfaden.....	57
3.4.1 Konzept der Lehrpfade.....	57
3.4.2 Beispiel eines WiBA-Net Lehrpfads.....	60
3.5 Entwicklung einer Software zur Unterstützung von Präsenzveranstaltungen....	61
3.6 Einsatz des WiBA-Net zur Vorlesungsunterstützung.....	63
4 Aufbereitung des Textkorpus.....	65
4.1 Notwendigkeit der Aufbereitung des Textkorpus.....	65
4.2 Konvertierung der Wissenseinheiten.....	67
4.2.1 Auslesen der Inhalte.....	67
4.2.2 Normierung der Inhalte.....	69
4.2.3 Rückführung von Worten auf ihre Stämme.....	71
4.2.4 Kompositazerlegung.....	73
4.3 Textstatistische Berechnungen.....	75
4.3.1 Bestimmung der Kookkurenzen.....	75
4.3.2 Vereinigung von Satz- und Nachbarkookkurenzen.....	76
4.3.3 Bestimmung der Ähnlichkeit von Begriffen.....	78
4.4 Datenspeicherung.....	79
4.5 Analyse der Inhalte des WiBA-Net.....	81
4.6 Erweiterung des Systems.....	85
5 Automatische Generierung von Lehrpfaden.....	89

5.1 Grundlagen der Lehrpfaderstellung.....	89
5.1.1 Bedeutung für das didaktische Konzept des WiBA-Net.....	89
5.1.2 Aufbau des Systems.....	91
5.1.3 Ablauf des Algorithmus.....	93
5.1.3.1 Schritte zur Lehrpfaderstellung.....	93
5.1.3.2 Länge des Lehrpfads.....	94
5.1.3.3 Anzeige der Lehrpfade.....	97
5.2 Suchanfrage.....	98
5.2.1 Suchtermerweiterung.....	98
5.2.2 Suchergebnisse	100
5.2.3 Eingangsgrößen für die Relevanzberechnung.....	100
5.3 Relevanz der Seiten	102
5.3.1 Zugehörigkeit von Wissensseinheiten zum Lehrpfad	102
5.3.2 Fuzzyfizierung.....	104
5.3.3 Inferenz	108
5.3.4 Defuzzyfizierung.....	115
5.3.5 Berücksichtigung von Unterthemen.....	116
5.3.6 Weiterführende Seiten.....	120
5.4 Anordnung der Lehrpfadseiten.....	121
5.4.1 Vorüberlegungen.....	121
5.4.2 Clustering der Wissensseinheiten.....	122
5.4.3 Sortierung der Cluster.....	125
5.4.3.1 Ordnungsprinzipien.....	125
5.4.3.2 Bestimmung der Kennwerte.....	129
5.4.3.3 Bestimmung der Varianten.....	130
5.4.4 Hierarchische Gliederung.....	133
6 Ergebnisse der Lehrpfaderstellung.....	136
6.1 Lehrpfade als Ergebnis von Suchanfragen.....	136
6.2 Analyse automatisch erstellter Lehrpfade.....	137
6.3 Lehrpfad 1.....	137
6.4 Lehrpfad 2.....	150
6.5 Lehrpfad 3.....	155

7 Zusammenhang zwischen Suchanfrage und Lehrpfadqualität.....	160
7.1 Notwendigkeit und Möglichkeiten der Bewertung.....	160
7.2 Relevanz der Lehrpfade bezüglich Suchanfragen.....	162
7.3 Relevanz der Lehrpfade unter Berücksichtigung lernzielähnlicher Seiten.....	173
7.4 Schlussfolgerungen.....	174
8 Zusammenfassung und Ausblick.....	177
8.1 Zusammenfassung der Ergebnisse.....	177
8.2 Didaktischer Mehrwert des Systems.....	177
8.3 Ausblick.....	178
9 Literaturverzeichnis.....	180
Anhang A: Verzeichnis verwendeter Formelzeichen.....	187
Anhang B: Glossar wichtiger Fachbegriffe.....	190
Anhang C: Wortendungen.....	191
Anhang D: Ausschlussworte.....	192
Anhang E: Im Projekt WiBA-Net integrierte Hochschullehrer.....	193
Anhang F: Liste der Wissensseinheiten des WiBA-Net.....	194
Anhang G: Ausgewählte Seiten des Lehrpfads „Konsistenzbestimmung“.....	208
Anhang H: Quelltext einer Wissensseinheit und ihre Anzeige im Browser.....	214
Anhang I: Seiten im automatisch generierten Lehrpfad.....	217
Anhang J: Ausgewertete Lehrpfade	224
Anhang K: Regressionsrechnung zu den untersuchten Lehrpfaden.....	243

1 Motivation und Zielsetzung

1.1 Ausgangssituation

E-Learning an Hochschulen

In den letzten Jahren hat sich E-Learning in der universitären Ausbildung mehr und mehr durchsetzen können. Hochschullehrer¹ aus allen Fachrichtungen verwenden heute E-Learning. Die Spannweite dessen, was als E-Learning bezeichnet wird, reicht dabei von ins Internet gestellten PDF-Skripten bis zu komplett online abgewickelten Seminaren. Zu verdanken ist die Entwicklung des E-Learning an den Hochschulen sowohl engagierten Professoren und ihren Mitarbeitern, die schon früh in Leuchtturmprojekten die Potentiale dieser Lehr- und Lernform aufzeigten, als auch einer umfangreichen Anschubfinanzierung durch Bund und Länder, z.B. im Rahmen des BMBF-Förderprogramms „Neue Medien in der Bildung“ [KINDT 2006]. Hierdurch wurde die Umsetzung von E-Learning an den Hochschulen in größerem Umfang möglich. Damit wurden die Vorteile einer durch E-Learning angereicherten Lehre in der Ausbildung sichtbar.

Der Einsatz und vor allem die Produktion von E-Learning sind mit nicht unerheblichem Aufwand verbunden. Um die Nachhaltigkeit der erfolgten Investitionen zu steigern, gibt es unterschiedliche Bemühungen zur Standardisierung der Inhalte. Die Standardisierung kann sowohl innerhalb einer Hochschule stattfinden, z.B. durch den Einsatz und die Förderung einer einzigen, zentralen Lernplattform, als auch auf einer (lehr-)institutionsübergreifenden Ebene, z.B. durch die Benutzung einheitlicher Standards für den Content oder die Metadaten. Im ersten Fall ist neben der Wiederverwendbarkeit der Inhalte an einer Hochschule das Ziel die Vereinheitlichung und damit Vereinfachung der Lernplattformbenutzung für die Studierenden. Im zweiten Fall wird die Standardisierung eher aus politischen bzw. wirtschaftlichen Gründen vorangetrieben [SCHULMEISTER 2001].

Jedoch finden sich vielerorts auch nicht-standardisierte Einzelprojekte, teilweise sind auch die unterschiedlichen Standards genügenden E-Learning-Inhalte verschiedener Hochschulen nicht ohne weiteren Aufwand miteinander kompatibel. Eine nachhaltige Nutzung von Inhalten wird dadurch erschwert. Zusätzlich kann der Fall auftreten, dass Lehrende bevorzugt eigenes Lehrmaterial einsetzen wollen und deswegen nicht auf bereits produzierte E-Learning-Inhalte zurückgreifen möchten. Die Aufbereitung eigenen Materials für den E-Learning-Einsatz ist jedoch aufwändig.

¹ Wegen der besseren Lesbarkeit wird bei allen Bezeichnungen für Personen in dieser Arbeit die männliche Form gewählt. Selbstverständlich sind immer sowohl die männliche als auch die weibliche Form gemeint.

Bologna-Prozess

Parallel zur Verstetigung des E-Learnings an deutschen Hochschulen erfolgt die Umsetzung des Bologna-Prozesses, die Einführung eines zweistufigen Systems von Studienabschlüssen (Bachelor und Master). Dabei werden auch an deutschen Hochschulen modularisierte Studiengänge eingeführt. Neben einer Steigerung der Mobilität der Studierenden ist das Ziel des Prozesses auch eine Verkürzung der Studiendauer [SCHWARZENBERGER 2005]. Die Vergleichbarkeit der Studienleistungen zwischen verschiedenen Universitäten im In- und Ausland wird durch das „European Credit Transfer and Accumulation System“ (ECTS) mittels der Vergabe von allgemein anerkannten „Credit Points“ (CP) sichergestellt [EUROPÄISCHE BILDUNGSMINISTER 1999]. Die Vergabe orientiert sich an Kriterien, die über die Anzahl der Semesterwochenstunden – also die Dauer der Präsenzveranstaltungen – hinausgehen. Damit wird der für die Studierenden veranschlagte so genannte „Workload“, also die Arbeitsbelastung, sichtbar [DIRECTORATE-GENERAL 2005]. So sind in die von den Studierenden voraussichtlich benötigten Arbeitszeiten neben der Dauer der eigentlichen Lehrveranstaltungen auch die Zeiten für Vor- und Nachbereitung dieser Veranstaltungen inkl. Literaturrecherche und Prüfungsvorbereitung einzurechnen. Analog dazu kann eine wöchentlich stattfindende Lehrveranstaltung als aus drei Phasen bestehend betrachtet werden: Vorbereitung auf die Präsenzveranstaltung im Selbststudium, Teilnahme an der Präsenzveranstaltung und anschließend die Nachbereitung der Veranstaltung, wiederum im Selbststudium. Die Verkürzung der Studiendauer, einhergehend mit unmittelbar nach jedem Semester stattfindenden Prüfungen, bewirkt eine höhere Lernbelastung der Studierenden. Die von STEINBORN 2007 festgestellte Tendenz, dass Studierende sich vornehmlich in einer Prüfung verlangtes Wissen aneignen, anstatt sich mit Hintergründen auseinander zu setzen, dürfte sich damit noch verstärken. Eine teilweise Entlastung der Studierenden kann durch sinnvoll eingesetzte E-Learning-Elemente erreicht werden, die den Nutzern eine freiere Zeiteinteilung erlauben. Die Möglichkeit zum Wiederholen und Nachschlagen des Lernstoffs, auch in den höheren Semestern, kann den Studierenden durch eine solche Lösung ebenfalls angeboten werden.

Lernnetzwerk WiBA-Net

Die vorliegende Arbeit baut auf den Ergebnissen des Forschungsprojekts WiBA-Net (Lernnetzwerk „Werkstoffe im Bauwesen“) [GRÜBL ET AL. 2004, vgl. auch Kapitel 3] auf. WiBA-Net ist ergänzend zu bestehenden Präsenzvorlesungen konzipiert. Es deckt inhaltlich das komplette Themengebiet der „Werkstoffe im Bauwesen“ ab, wie es von den dieses Fach an deutschsprachigen Hochschulen Lehrenden im Memorandum der Hochschullehrer [REINHARDT 2000, vgl. Abschnitt 3.2] als Mindestlehrinhalt festgelegt wurde.

Neben der Erstellung der Inhalte wurden auch didaktische Konzepte für den Einsatz des Lernnetzwerks im Rahmen der Vorlesung erarbeitet. Die E-Learning-Inhalte können dabei sowohl in den Präsenzveranstaltungen als auch zum Selbstlernen genutzt werden. Für das Selbststudium stehen die Inhalte den Lernern in Form der im Rahmen des Projekts entwickelten „Lehrpfade“ zur Verfügung. Dies sind Lerneinheiten, die aus Inhaltsseiten, die auch multimediale Elemente enthalten können, bestehen. Die Anordnung dieser Seiten zum Lehrpfad erfolgt nach bestimmten didaktischen Kriterien (vgl. Abschnitt 3.4.1).

1.2 Verbesserungspotential beim Einsatz von E-Learning

Vor- und Nachteile von E-Learning

Der Mehrwert von richtig eingesetztem E-Learning liegt in einer gesteigerten Effizienz, sowohl in ökonomischer als auch in didaktischer Hinsicht. So können komplexe oder wenig anschauliche Inhalte durch eine multimediale Darstellung leichter vermittelt werden. Einmal erstellte Inhalte können, auch in neuer Zusammenstellung, wiederverwendet werden. Zudem sparen Lerner Zeit und Kosten, wenn Inhalte online zugreifbar sind und keine persönliche Teilnahme an einer Präsenzveranstaltung notwendig ist. Der Mehrwert zeitlicher und räumlicher Flexibilität beim E-Learning ist ökonomisch sicherlich in der Weiterbildung am besten messbar [BREMER 2004], die Effizienzsteigerung durch mehr Flexibilität ist jedoch auch auf Studierende an Hochschulen übertragbar. Schließlich eröffnet E-Learning neue didaktische Möglichkeiten, denn es erlaubt eine innovative Art der Integration von Präsenzlehre und Selbststudium. Konkret messbar werden die Vorteile des E-Learning, wenn man die Prüfungsergebnisse von Nutzern eines E-Learning-Systems und Studierenden, die dieses nicht genutzt haben, vergleicht. WEHLING 2008 stellt hier z.B. bei Studierenden in den Fächern „Werkstoffe im Bauwesen“ und „Bauphysik“ eine signifikante Verbesserung der Notenstruktur bei den E-Learning-Nutzern fest.

Demgegenüber stehen die Nachteile des E-Learning: Allgemein kann festgestellt werden, dass es sich dabei nicht um ein „Allheilmittel“, keinen „elektronischen Nürnberger Trichter“ [GRÜBL ET AL. 2003] handelt; vielmehr sollte E-Learning nur da eingesetzt werden, wo es didaktisch sinnvoll ist [KERRES 2001]. Des Weiteren ist die erstmalige Produktion von E-Learning-Inhalten teuer – der Arbeitsaufwand für die Herstellung und Umsetzung von Inhalten kann leicht hundertfach über der späteren Lernzeit liegen. Selbst ohne Videos und Animationen kann eine Stunde E-Learning-Unterricht also schnell mehrere zehntausend Euro kosten [KERRES 2001]. Ferner weist E-Learning Akzeptanzprobleme sowohl bei Lehrenden als auch bei Lernenden auf. Diese Probleme können zum einen in der Lehr- bzw. Lernweise des Einzelnen begründet sein: Dasselbe Lernangebot wird z.B. nicht von allen Studierenden gleich gut angenommen [APEL 1999]. Zum anderen können Ak-

zeptanzprobleme technischer Natur sein. Dies trifft sowohl auf Lehrende wie auch auf Lernende zu. Studierende schätzen ihre Computerkenntnisse im Allgemeinen zwar selbst als relativ gut ein, jedoch ist ihre tatsächliche Fähigkeit, die Informationsquelle Internet kritisch und effektiv zu nutzen zu hinterfragen [WEHLING 2008] [STEINBORN 2007] [WHITMIRE 2004]. Die Konsequenz ist, dass Studierende explizit in der Anwendung von Internetrecherchen und E-Learning geschult werden sollten [STEINBORN 2007], um eine sinnvolle Anwendung dieses Hilfsmittels zu ermöglichen und dessen Akzeptanz zu erhöhen. Wahrscheinlich wird eine volle Akzeptanz von E-Learning als gleichberechtigte Art zu lernen sowohl bei Studierenden als auch bei Lehrenden noch eine gewisse Zeit benötigen.

Für die Lehrenden als Autoren von E-Learning-Inhalten liegen verschiedene Hilfsmittel vor – von HTML-Editoren über Animationsprogramme bis hin zu Lernplattformen. Jedoch ist die Bedienung dieser Programme teilweise komplex und die Einarbeitung zeitaufwändig und schwierig. Hinzu kommt noch der Aufwand für die Auszeichnung erstellter Inhalte mit geeigneten Metainformationen, wie z.B. Schlagworten, bzw. für eine Kategorisierung der Inhalte. Beides wird durchgeführt, um die Wiederverwendbarkeit der Inhalte zu erhöhen. Für die Nachhaltigkeit eines E-Learning-Systems ist aber die Benutzbarkeit für dessen Autoren von großer Bedeutung [SESINK 2003], dies gilt im Rahmen der universitären Ausbildung insbesondere für die beteiligten Hochschullehrer. Die Erstellung und Anzeige der Lehrpfade erfolgt im Projekt WiBA-Net mit dem Programm *SAP Learning Solution*². Die Seiten der Lehrpfade lassen sich hier verschiedenen Kategorien zuordnen, was das Sortieren der Lehrpfade nach verschiedenen didaktischen Kriterien ermöglicht. Die Zusammenstellung der Inhalte und ihre spätere Pflege sind mit diesem Programm jedoch sehr aufwändig. Nicht zu vernachlässigen ist schließlich der soziale Faktor von Präsenzlehrveranstaltungen, also der persönliche Kontakt der Lernenden zum Lehrenden und untereinander (vgl. hier und im Folgenden [APEL 1999]). Hochschullehrer können in einer kommunikativen Lehr-Lern-Situation die Zuhörenden aktivieren: Die Präsenzlehre lebt insbesondere von spontaner Darstellung der Inhalte, die persönliche Begeisterung eines Dozenten für sein Fachgebiet kann zu einer besseren Motivation der Lernenden beitragen. Studierende, gerade im Grundstudium, bringen häufig auch nicht die notwendige Selbstdisziplin für ein völlig eigenständiges Lernen von E-Learning-Inhalten auf, ein gewisser Anwesenheitszwang in der Vorlesung ist hier von Vorteil. Auch eine obligatorische Bearbeitung bestimmter E-Learning-Inhalte kann insofern sinnvoll sein, um die Studierenden an die Arbeit mit einem E-Learning-System zu gewöhnen.

2 Das Programm wurde zuvor unter dem Namen „L³“ vermarktet.

E-Learning bei verschiedenen Arten von Lehrveranstaltungen

Die an Hochschulen angebotenen Lehrveranstaltungen lassen sich grundsätzlich verschiedenen Typen zuordnen, die unterschiedlichen didaktischen Ansprüchen genügen und ein unterschiedliches Vorgehen des jeweils Lehrenden erfordern, beispielsweise Vorlesungen, Seminare, Übungen, Praktika oder Projektseminare [vgl. z.B. VORLESUNGSVERZEICHNIS TUD 2008]. E-Learning-Elemente werden heute vor allem im Rahmen von Seminaren bzw. Projektseminaren (Koordination und Kommunikation der Teilnehmer, Abgabe der Seminararbeiten, Publikation der Ergebnisse, vgl. z.B. [MENZEL U. POTTGIESSER 2003], [HÄFELE U. MAIER-HÄFELE 2004]) und Praktika (Vorbereiten oder Ersetzen von Präsenzpraktika durch Animationen und Simulationen, vgl. z.B. [HAGEMANN 2003], [RÖHRIG U. JOCHHEIM 1999]) eingesetzt. Entsprechende Software und didaktische Konzepte existieren.

Einige Ansätze zur Einführung von E-Learning beim Lehrveranstaltungstyp Vorlesung verfolgen den Weg des Ersetzens eines Teils der Präsenzveranstaltungen durch E-Learning als eine Umsetzung des Konzepts des Blended-Learning³. Auch die vollständige Durchführung von Vorlesungen mittels E-Learning ist möglich. So zeigen STEINBORN U. LANGE 2007 ein Beispiel für die Durchführung einer hochschulübergreifenden Lehrveranstaltung, an der Studierende fremder Hochschulen über das Internet teilnehmen können.

Häufig wird bei universitären Lehrveranstaltungen eine Ergänzung der Präsenzlehre durch E-Learning-Elemente dem vollständigen Ersatz vorgezogen [APEL 1999]. Auf diese Weise kann man die didaktischen Vorteile, die Präsenzlehre und virtuelle Lehre bieten, optimal nutzen. Beispielsweise können Grundlagen und Faktenwissen in Form von E-Learning im Selbststudium erarbeitet werden. In der Präsenzveranstaltung kann dann zur vertiefenden Darstellung der Inhalte und ihrer Diskussion übergegangen werden [GRÜBL ET AL. 2003] [APEL 1999]. Die Inhalte der Präsenzveranstaltung können dann wiederum zur Nachbereitung online veröffentlicht werden.

Vorlesungen werden heute oftmals nur durch die Bereitstellung von Skripten und anderen Unterlagen im Internet „auf E-Learning getrimmt“. Dies ist zwar eine Verbesserung gegenüber dem Kopieren und Verkaufen derselben Unterlagen, stellt jedoch je nach Definition noch kein echtes E-Learning dar. Dies wäre dann der Fall, wenn ein didaktischer Mehrwert entsteht. Sinnvoll kann hier z.B. der Einsatz von Multimedia-Elementen auch in Präsenzveranstaltungen sein, sodass ein Wiedererkennungswert mit angebotenen Selbstlernmaterialien zur Vor- und Nachbereitung entsteht. Studierenden ist daneben auch die Identifikation des Lehrenden mit dem eingesetzten E-Learning-System wichtig [SESINK 2003].

3 zur Definition vgl. Abschnitt 2.1.4

Verbesserungspotential

In den derzeitigen Nachteilen des E-Learnings liegt gleichzeitig das Verbesserungspotential für E-Learning-Systeme: Bereitstellung von Werkzeugen für bislang vernachlässigte didaktische Einsatzmöglichkeiten, Effizienzsteigerung in der Produktion von Lernelementen, einhergehend mit leichter Benutzbarkeit, die Stärkung der sozialen Komponenten einer Vorlesungsreihe im Rahmen des so genannten „Blended Learning“, einem Wechsel zwischen E-Learning und Präsenzeinheiten, sowie die Erhöhung der Akzeptanz von E-Learning durch das Angebot selbst bestimmten Lernens.

Ein neuer Ansatz der Medienpädagogik ist es, die Diversität der Lernenden bei der Planung von E-Learning-Angeboten zu berücksichtigen [GRÜBL ET AL. 2003] [SCHULMEISTER 2004]. Dieser Ansatz lässt sich sinngemäß auch auf die unterschiedlichen Lehrweisen verschiedener Hochschullehrer übertragen.

1.3 Ziele und Vorgehensweise

In der vorliegenden Arbeit wird ein Vorschlag zur automatischen Generierung von individuellen Lerneinheiten („Lehrpfade“) aus textbasierten Wissensseinheiten gemacht. Der Vorschlag besteht aus der didaktischen und technischen Konzeption und der Implementierung eines prototypischen Programmsystems. Dabei sollen sowohl Lernende als auch Lehrende berücksichtigt werden. Bei den Anwendergruppen sollen neue Freiheitsgrade beim Wissenserwerb bzw. bei der Wissensvermittlung eingeräumt werden.

Von Studierenden automatisch generierte Lehrpfade sollen sich direkt zum Lernen und Nachschlagen nutzen lassen.

Lernende werden durch speziell für ihre Anfrage erstellte Lehrpfade bei der Befriedigung ihres individuellen Informationsbedürfnisses unterstützt und können nach eigenen Interessen lernen. Auch die Wiederholung ausgewählter Inhalte wird so ermöglicht. Dabei sollen die Lerner nach dem Durcharbeiten der zum Lehrpfad gehörenden Wissensseinheiten in der Lage sein, die gestellte Frage selbst zu beantworten und zukünftig eigene Entscheidungen auf der Basis des erworbenen Wissens treffen zu können. Prinzip des Systems ist, dass keine wissenschaftlich strenge Antwort auf eine gestellte Frage gegeben wird (wie bei einem „Expertensystem“), sondern dass eine geeignete Teilmenge der Wissensseinheiten dem Nutzer als didaktisch sinnvoller und inhaltlich strukturierter Lehrpfad präsentiert wird. Der Unterschied des vorgeschlagenen Systems zu einer Internet-Suchmaschine ist, dass im vorliegenden Fall nicht einzelne Treffer nach einer bestimmten Reihenfolge angezeigt werden. Vielmehr wird die Menge der relevantesten Treffer ausgewählt. Alle ande-

ren Treffer werden verworfen. Die ausgewählte Menge wird dem Nutzer als Ganzes, jedoch in einer Abfolge und in strukturierter Form, angezeigt.

Auch Hochschullehrer sollen das System nutzen können, um auf der Basis automatisch erstellter Lehrpfade schnell E-Learning-Inhalte produzieren zu können. Sie sollen diese Inhalte aber nach individuellen fachlichen Schwerpunkten und didaktischen Präferenzen manuell überarbeiten und ergänzen können.

Die Produktion von E-Learning-Inhalten kann, wie bereits angesprochen, sehr aufwändig sein. Vorhandene E-Learning-Lerneinheiten passen jedoch manchmal nicht vollständig in das inhaltliche und didaktische Konzept eines Dozenten für eine bestimmte Präsenzveranstaltung. Dies gilt insbesondere, wenn der Autor der Inhalte und der Dozent unterschiedliche Personen sind. Mit dem vorgeschlagenen System kann jeder Dozent nach seinem Ermessen aus zur Verfügung stehenden Textmaterialien neue Lerneinheiten für Studierende erstellen. Er ist dabei frei in der Auswahl und in der Anordnung der einzelnen Inhalte.

Die Zusammenstellung der Lerneinheiten soll ausschließlich auf der Basis der in einem E-Learning-System vorhandenen Texte und aus diesen extrahierbaren Informationen ablaufen.

Die textlichen Inhalte des E-Learning-Systems sollen weitgehend automatisch aufbereitet und anhand textstatistischer Kriterien ausgewertet werden. Dies geschieht u.a. durch den Einsatz von Methoden der Fuzzy Logic. Die für die Lehrpfade relevanten Wissensseinheiten werden mit Information-Retrieval-Methoden ausgewählt und dann in eine angemessene Reihenfolge gebracht. Sie umfassen das Themengebiet der Anfrage, soweit es von den Inhalten des Systems abgedeckt ist. Die Autoren ersparen sich so eine manuelle Vernetzung der Inhalte zusätzlich zum Erstellen der Wissensseinheiten. Ebenso kann weitgehend auf eine Metadatenvergabe für textbasierte Wissensseinheiten verzichtet werden.

Das beschriebene System soll als lauffähiger Prototyp implementiert werden, mit dem Lehrpfade nach dem beschriebenen Verfahren erstellt werden können.

So kann die Funktionsfähigkeit der einzelnen Teile des aus dem Vorschlag resultierenden Algorithmus und des Zusammenwirkens dieser Einzelteile sichergestellt werden. Ferner können etwaige Probleme, insbesondere das Zusammenwirken der Einzelaspekte betreffend, erkannt werden.

Das beschriebene System wird am Beispiel der Inhalte des Lehrfachs „Werkstoffe im Bauwesen“, die für das Lernnetzwerk „WiBA-Net“ erstellt wurden, dargestellt.

Die Inhalte dieses Lernnetzwerks decken das Fach „Werkstoffe im Bauwesen“ inhaltlich vollständig ab. Es können also mit dem vorgeschlagenen System zu Suchanfragen aus diesem Themengebiet prinzipiell inhaltlich korrekte Lehrpfade erstellt werden. Das WiBA-Net wird in der Praxis eingesetzt. Sein eigens entwickeltes didaktisches Konzept wurde erprobt und kann als Grundlage einer Erweiterung dienen, die sich durch die Möglichkeiten des vorgeschlagenen Systems ergibt. Die Übertragbarkeit auf andere textbasierte E-Learning-Inhalte ist gegeben, da die notwendige Aufbereitung und Auswertung der Inhaltsdateien beschrieben wird und weitgehend automatisch abläuft. Das für die Auswahl von Seiten für konkrete Lehrpfade verwendete Inferenzmodul enthält keine auf die speziellen verwendeten Inhalte bezogenen Regeln, sodass es für die Nutzung anderer Inhalte nicht umgeschrieben oder ergänzt werden muss.

Für das System soll abschließend eine statistische Methode entwickelt werden, mit der eine Prognose über die voraussichtliche Qualität eines automatisch generierten Lehrpfads auf der Basis der verwendeten Suchbegriffe abgegeben werden kann.

Dies ist notwendig, da automatisch erzeugte Lehrpfade prinzipiell auch irrelevante Seiten enthalten können. Insbesondere Lernenden mit geringen Vorkenntnissen soll so die Möglichkeit gegeben werden, die inhaltliche Qualität des von ihnen erzeugten Lehrpfads zu beurteilen. Die Methode basiert auf den textstatistischen Eigenschaften der Suchbegriffe.

1.4 Aufbau der Arbeit

Kapitel 2 erläutert die Grundlagen des E-Learnings und dessen didaktische Theorien. Darüber hinaus werden Methoden des Information-Retrieval vorgestellt und die Grundlagen von Sprachwissenschaft und Computerlinguistik erklärt.

Kapitel 3 beschreibt das Lernnetzwerk WiBA-Net, auf dessen Grundlage diese Arbeit entstanden ist und dessen Inhalte zur prototypischen Implementierung des Systems zur automatischen Lehrpfaderzeugung herangezogen wurden.

Kapitel 4 beschreibt die Aufbereitung der Inhalte einer Wissensdomäne für den Einsatz textstatistischer und computerlinguistischer Verfahren anhand des WiBA-Net.

Kapitel 5 erläutert die Vorgehensweise bei der Erzeugung eines Lehrpfads, der die in einer Suchanfrage implizierten Themenbereiche abdeckt.

Kapitel 6 analysiert die Ergebnisse der Lehrpfaderzeugung anhand exemplarischer Lehrpfade zu drei verschiedenen Themen.

Kapitel 7 umfasst die Untersuchung von automatisch generierten Lehrpfaden hinsichtlich ihrer inhaltlichen Qualität und leitet daraus eine Methode der Vorhersage der Lehrpfadqualität anhand der Eigenschaften der Suchanfrage ab.

Kapitel 8 fasst die Arbeit zusammen und liefert Ansatzpunkte für weiterführende Arbeiten.

Ein Verzeichnis verwendeter Formelzeichen findet sich in Anhang A

Ein Glossar wichtiger Fachbegriffe findet sich in Anhang B.

2 Stand des Wissens

2.1 E-Learning

2.1.1 Was ist E-Learning?

Zum Begriff des E-Learning existiert keine allgemein akzeptierte Definition. H. Dichanz versteht unter E-Learning „ein Informations- und Lehrsystem, das Lernern auf elektronischem Wege Informationen, Navigationshilfen, Placementhinweise (Selbsttests) und Anknüpfungspunkte an jedem Ort und zu jeder Zeit in unterschiedlicher Form individuell oder für Gruppen zur Verfügung stellt“ [DICHANZ 2005]. Er stellt damit konkrete Anforderungen an die Funktionalitäten eines solchen Systems. M. Kerres hingegen bezeichnet E-Learning als den „Oberbegriff für alle Varianten internetbasierter Lehr- und Lernangebote“ [KERRES 2001]. L. Abicht definiert den Kern des E-Learning ohne Bezug zur Technik und nennt das „selbstständige informelle (entpädagogisierte) Lernen“ eine der „Säulen“ des E-Learning [ABICHT U. DUBIEL 2003].

Die Definition von Kerres ist umfassend und schränkt den Begriff nicht ein, indem bestimmte Funktionalitäten gefordert werden. Man sollte jedoch berücksichtigen, dass durch diese Definition auch einige Internetanwendungen einbezogen werden, die gemeinhin nicht als E-Learning betrachtet werden: als Hilfsmittel z.B. Foren und Blogs, die zur Informationsrecherche herangezogen werden können, sowie als elektronische Medien Textdokumente mit fachlichen Inhalten, universitäre Skripte im PDF-Format oder elektronisch publizierte Dokumente, die über die Webseiten von Bibliotheken zugänglich sind.

Will man diese Anwendungen elektronischer Medien und Hilfsmittel aus der Definition von E-Learning ausschließen, so kann man E-Learning als das Lehren und Lernen mit Hilfe digitaler Medien und elektronischer Hilfsmittel betrachten, wenn dabei durch den gleichzeitigen Einsatz dieser Medien und Hilfsmittel ein didaktischer Mehrwert bei den Lernern entsteht.

2.1.2 Pädagogisch-didaktische Aspekte und Anwendungsbeispiele aus dem E-Learning

Man unterscheidet in der Didaktik im Allgemeinen drei Paradigmen („*Lerntheorien*“). Dies sind der Behaviorismus, der Kognitivismus und der Konstruktivismus (vgl. Tabelle 2.1). Die drei Ansätze lösten einander im Verlauf der letzten Jahrzehnte als vorherrschende Theorien ab, auf deren Grundlage versucht wurde, Wissensvermittlung didaktisch zu gestalten. Der Behaviorismus war in den 1950er und 60er Jahren am populärsten, wurde dann in den 70er und 80er Jahren vom Kogni-

tivismus als dominanter Lerntheorie abgelöst und dieser wiederum ab Mitte der 90er Jahre vom Konstruktivismus. Jedoch finden sich auch in den späteren Phasen noch große und auch berechnete Anteile der vormals moderneren Theorien. So bauen auch heute noch z.B. Programme zum Erlernen von Fremdsprachen auf behavioristischen Ansätzen auf. Auch wird heute aus der Sicht der Mediendidaktik keiner der Ansätze als überlegen angesehen [KERRES 2001], sondern es wird empfohlen, pragmatisch zu versuchen, die jeweils sinnvollen und am besten für die jeweilige Lehr-Lern-Aufgabe passenden Anteile zu identifizieren und zu integrieren. Ebenso wenig kann es selbst für ein Studienfach eine einheitliche didaktische Methode geben, sondern es werden „für verschiedene Lernziele [...] spezifische Angebote benötigt“ [STEINEBACH 2005].

Tabelle 2.1: Gegenüberstellung der Lernparadigmen

Lernparadigma	Theorie bzgl. Wissen	Beispiele
Behaviorismus	Wissen wird im Gehirn „gelagert“ [STEINEBACH 2005].	Programmierter Unterricht, Sprachlabore
Kognitivismus	Wissen wird vom Lerner verarbeitet.	Intelligente Tutoring Systeme
Konstruktivismus	Wissen wird konstruiert.	Projektaufgaben

Der Behaviorismus wurzelt in der Verhaltensforschung und wurde von B. F. Skinner zu einem lernpsychologischen Ansatz entwickelt. Er hat im E-Learning bis heute einen starken Einfluss, da er bei frühen E-Learning-Anwendungen sehr verbreitet war [KERRES 2001]. Als Fachrichtung der Psychologie besagt er, dass Verhaltensweisen (Responses – R) durch Reize (Stimuli – S) ausgelöst werden. Ein bekanntes Beispiel hierfür ist der Pawlowsche Hund, der mit Speichelproduktion (R) auf das Riechen von Futter (S) reagiert und sich durch gleichzeitiges Klingeln einer Glocke dahingehend konditionieren lässt, dass später allein der Klang der Glocke ausreicht, um die Speichelproduktion anzuregen.

Von Skinner wurde hieraus das so genannte „operante Konditionieren“ erarbeitet, davon ausgehend, dass ein Verhalten nicht nur reagierend sondern auch spontan auftreten kann und dann durch „Verstärkung“ (z.B. eine Belohnung) konditioniert werden kann [SCHULMEISTER 2002]. Eine Anwendung der operanten Konditionierung ist der „Programmierte Unterricht“, eine Lehrform, in der der zu vermittelnde Stoff in kleinste Einheiten geteilt wird, die nacheinander präsentiert werden und deren Verständnis unmittelbar überprüft wird (Stimulus-Response-Ketten). Diese Vorgehensweise scheint für ein computergestütztes Lernen prädestiniert, da die Verstärkungsmechanismen hier konsequenter und effektiver eingesetzt werden können [KERRES 2001]. Die Verstärkung / Belohnung für die richtige Antwort auf eine Frage ist die Bearbeitung der nächsten Einheit. Bei falschen Antworten ignoriert das System den Lerner und wiederholt die Frage oder springt im Stoff

zurück. Die Rate der Verstärkungen ist dabei idealerweise dem Leistungsniveau anzupassen, so dass der Lerner sich für eine weitere Belohnung mehr anstrengen muss („intermittierende Verstärkung“) [KERRES 2001]. Der „Programmierte Unterricht“ konnte sich jedoch nicht durchsetzen. Heute werden Programme nach dem behavioristischen Ansatz z.B. zum Einüben von Faktenwissen („Drill & Practice“) verwendet [ABICHT U. DUBIEL 2003].

Kritik am Behaviorismus bezieht sich vor allem auf das Problem, dass beim Lernen nicht „die richtigen Reaktionen auf bestimmte Fragen“ [KERRES 2001] eingeübt werden sollen. Vielmehr sollte das gelernte Wissen im Kontext abrufbar sein und damit auch Antworten auf neue Fragen ermöglichen. Weiterhin behindert die Teilung des zu lernenden Stoffes in kleinste Einheiten „das Verständnis des Ganzen“ [STEINEBACH 2005].

Während im Behaviorismus die Vorgänge im Gehirn des Lernalers nicht betrachtet werden, da es nur auf beobachtbare Reaktionen auf Reize ankommt, werden im Kognitivismus auch diese internen Prozesse berücksichtigt. Es wird hier angenommen, dass Wissen auf „adäquaten internen Verarbeitungsprozessen“ beruht, der Beitrag des Lernenden selbst zum Lehr-Lernprozess wird im Gegensatz zum Behaviorismus berücksichtigt [STEINEBACH 2005]. Neu aufgenommene Informationen werden mit vorhandenem Wissen und im Kontext dieses vorhandenen Wissens kombiniert („Analyse durch Synthese“) [KERRES 2001]. Eine weitere Annahme des Kognitivismus ist, dass verschiedene Lerninhalte in unterschiedlichen Teilen des Gedächtnisses gespeichert und auch unterschiedliche Verarbeitungsschritte notwendig sind, um diese Inhalte dauerhaft zu speichern [KERRES 2001]. Dies führte zu einer Klassifizierung von zu lernendem Wissen, zunächst durch die „Taxonomie der Lernziele“ von B. Bloom [BLOOM 1956]. Die Taxonomie beginnt auf dem untersten Level, der Kenntnis von Fakten, und endet bei der selbstständigen Produktion und Evaluation von Wissen. Weitere Klassifikationen von Lehr- und Lernprozessen stammen von Anderson und Gagné (vgl. Tabelle 2.2). Die Bloomsche Taxonomie soll es Lehrenden ermöglichen, Lernkontrollen so durchzuführen, dass beurteilt werden kann, welche Levels von den Lernern jeweils beherrscht werden und wo noch Bedarf der Unterstützung besteht.

Nach kognitivistischen Ansätzen soll der Lernende bei der Bearbeitung von Aufgaben vom Lehrer in der Rolle eines Tutors [ABICHT U. DUBIEL 2003] beobachtet und unterstützt werden und externes Feedback erhalten [STEINEBACH 2005]. Dabei ist der Wissensstand des Lernalers zu berücksichtigen. Da Lerner keine homogene Gruppe sind, muss neben ihrem jeweiligen Wissensstand auch ihre individuelle Lernstrategie berücksichtigt werden. Man spricht in diesem Zusammenhang von der „Diversität“ der Lerner⁴.

4 Die Diversität der Lerner hat neben den Lernstrategien noch andere Einflussfaktoren, u.a. kulturelle Faktoren oder die Motivation der Lerner, vgl. [SCHULMEISTER 2004].

Tabelle 2.2: Klassifizierung von Lernzielen bzw. Wissenstypen, vgl. [KERRES 2001]

Urheber	Klassifizierung
B. Bloom	<i>Lernziele:</i> Wissen → Verstehen → Anwenden → Analyse → Synthese → Evaluation
J. R. Anderson	<i>Wissenstypen:</i> deklarativ → prozedural → kontextuell
R. Gagné	<i>Lernprozesse:</i> Begriffe → Regeln → Problemlösen

Im Bereich des computergestützten Lernens bedeutete dies die Einführung von „adaptiven Systemen“ (z.B. Intelligente Tutoring Systeme, ITS), die sich auf den Lernfortschritt einstellen konnten. Solche Systeme registrieren und bewerten das Lernverhalten. Es erfolgt ein Vergleich mit einem Expertenmodell, die ermittelte Differenz zwischen dem Modell des Lerners und dem des Experten ergibt die Lerndefizite. Auf Basis dieser Defizite erfolgt dann die Auswahl der weiteren Lerneinheiten [KERRES 2001]. Ferner wird versucht, aus der Art eines gemachten Fehlers auf dessen Ursache und die entsprechenden Verständnisprobleme beim Lerner zu schließen. Problematisch bei adaptiven Systemen ist jedoch [KERRES 2001] [SCHULMEISTER 2004]:

- dass nur ungenau auf die Ursache des Fehlers eines Lerners geschlossen werden kann;
- dass sie domänenspezifisch sind, also die Inhalte eines ITS nicht einfach gegen andere Inhalte ausgetauscht werden können;
- dass die Vielzahl der zu berücksichtigenden Variablen und ihrer Interaktionen zu groß wird, um ein sinnvolles System implementieren zu können.

Aus kognitivistischen Ansätzen wurde von Gagné und anderen das Modell der Instruktionsereignisse⁵ entwickelt. Ein Lernprozess kann als Sequenz dieser Instruktionsereignisse verstanden werden [SCHULMEISTER 2002]. Unterschiedliche Lehrziele bedingen eine unterschiedliche Ausgestaltung der einzelnen Instruktionsereignisse, ihre Reihenfolge wird jedoch nicht verändert [KERRES 2001].

Zusammenfassend kann man festhalten, dass Lernen kein von außen gesteuerter Prozess ist, sondern auf jeden Fall „eine individuelle kognitive Handlung und aktive Leistung“ [DICHANZ 2005] eines Lerners erfordert. An dieser Stelle setzt der Konstruktivismus an.

Nach konstruktivistischen Ansätzen gilt Wissen „nicht mehr als Abbild der Wirklichkeit, sondern als eine individuelle Konstruktion des Menschen“ [STEINEBACH 2005] bzw. als „eine Funktion des Erkenntnisprozesses“ [SCHULMEISTER 2002]. Nach dem Konstruktivismus „existiert [Wissen] nicht

⁵ Die Instruktionsereignisse sind: 1. Aufmerksamkeit erlangen, 2. Lernziele nennen, 3. Wiederholen der Lernvoraussetzungen, 4. Präsentation von Informationen, 5. Anbieten von Lernhilfen, 6. Übung / Anwendung, 7. Rückmeldung, 8. Überprüfung des Lernerfolgs und 9. Transfer

unabhängig vom erkennenden Subjekt“, sondern wird – daher der Name – „im Akt des Erkennens konstruiert“. Dies führt zu der Schlussfolgerung, dass Wissen nicht „übermittelt“ werden kann, denn der Lerner muss es selber für sich konstruieren [SCHULMEISTER 2002]. Folglich sollte Lernen nach konstruktivistischen Ansätzen situationsbezogen („situierter“) stattfinden. Ein Modell für ein solches Lernen ist die so genannte „Kognitive Lehre“ [KERRES 2001], die auch als „Apprenticeship“ [STEINEBACH 2005] bezeichnet wird und die an die Handwerkslehre angelehnt ist. Hierbei werden die Lernenden durch Handeln in einer praxisnahen Umgebung von einem Experten in die jeweilige „Expertenkultur“ eingeführt.

Ein weiteres Modell ist das *geankerte Lernen* [KERRES 2001] [STEINEBACH 2005] („anchored instruction“ [SCHULMEISTER 2002]), bei dem in einer Lernsituation authentische Probleme dargestellt werden, um die Motivation und Aufmerksamkeit der Lernenden zu steigern. Geankertes Lernen findet auch im Rahmen des Lernens mit Projekten statt [KERRES 2001]. Für den Lernerfolg bei einem Projekt ist sowohl die „ganzheitliche“ Herangehensweise wichtig als auch die Komponente der Organisation der Lern- und Bearbeitungsabläufe.

2.1.3 Entwicklung des E-Learning

2.1.3.1 Historische Entwicklung

Als frühe Formen des E-Learnings in Deutschland kann man Funkkolleg und Telekolleg betrachten, die 1966 bzw. 1968 erstmals ausgestrahlt wurden und heute noch in modifizierter Form fortgesetzt werden (vgl. [BAYERISCHER RUNDFUNK 2004], [HESSISCHER RUNDFUNK 2004]). Ungefähr zeitgleich wurden hierzulande auch erstmals Computer zur Wissensvermittlung eingesetzt. Zunächst handelte es sich um die so genannte „Programmierte Instruktion“ nach dem behavioristischen Ansatz von B. F. Skinner [KERRES 2001]. Die Anfänge dieser Programme in den USA liegen in den 1950er Jahren. Bei der Programmierten Instruktion werden dem Lerner Fakten präsentiert und Fragen dazu gestellt („Drill & Practice“). Der Computer dient hierbei jedoch nur als reines Werkzeug zur effektiven Umsetzung der behavioristischen Theorie. Das Potential, den kognitiven Prozess zu verbessern, wird nicht genutzt. Die nächste Stufe stellte in den 1960er und 70er Jahren die „Computer Aided Instruction (CAI)“ dar [DELLA CORTE 2005]. Aus dem „Programmierten Unterricht“ entstanden die so genannten Autorensysteme, die es Lehrenden ermöglichen, selbst Lernprogramme zu erstellen. Dies war zunächst nur für die Methode des Programmierten Unterrichts möglich. Das System PLATO von 1960 fand erstmals größere Verbreitung [KERRES 2001]. Es erlaubte gegenüber dem System TICCIT aus den 1970er Jahren, das auf dem Instruktionsdesign basierte [SCHULMEISTER 2002], aber auf eine spezielle Methode festgelegt war, eine freiere

Erstellung von Lernprogrammen. PLATO hatte größeren Erfolg und spätere Autorensysteme orientierten sich an diesem Konzept [KERRES 2001].

Der Weg zum „Computer Assisted Learning (CAL)“ führte über die Entwicklung des Hypertexts Mitte der 1960er Jahre [DELLA CORTE 2005]. Hypertexte erlauben die nicht-lineare Exploration des Kontextes der zu lernenden Inhalte und setzten sich damit deutlich von den bisherigen vorgegebenen Lernwegen ab. In den 1970er und 80er Jahren wurden nach und nach besser werdende adaptive Systeme geschaffen. Zunächst leiteten die Programme aus der Beantwortung von Multiple-Choice-Fragen den weiteren Lernweg ab, später wurde versucht, den aktuellen Wissensstand bzw. den Lernfortschritt der Nutzer mit „Intelligenten tutoriellen Systemen (ITS)“ besser zu berücksichtigen, indem aus aktuellen Eingaben der Benutzer Rückschlüsse auf deren Kompetenz gezogen werden sollten [KERRES 2001]. Dieses Vorgehen setzte sich jedoch bislang nicht durch, da die Systeme die Art eines Fehlers und damit das eigentliche Verständnisproblem nicht zuverlässig erkennen. Gründe dafür liegen im unterschiedlichen Lernverhalten der Nutzer und in der dadurch entstehenden Vielzahl von zu berücksichtigenden Variablen sowie in der Schwierigkeit, einem erkannten Lernverhalten eine tatsächlich passende Lehrmethode zuzuordnen [SCHULMEISTER 2004].

Seit den 1990er Jahren werden in der Neuentwicklung von E-Learning-Systemen vermehrt konstruktivistische Ansätze verfolgt [KERRES 2001]. Dazu gehören der Einsatz von Medien als Werkzeuge zur Wissenskonstruktion in „Lernumgebungen“ [SCHULMEISTER 2004] oder das Lernen mit Projekten. Hierfür sind Systeme nötig, die die notwendige Kooperation zwischen Lernenden und Lehrenden ermöglichen und möglichst flexibel sind, um Lehrern und Lernern Gelegenheit zu einer möglichst natürlichen Arbeitsweise zu geben [DELLA CORTE 2005].

Auch das jeweilige „Transportmedium“ änderte sich im Laufe der Zeit: Frühe E-Learning-Applikationen, wie z.B. PLATO, liefen auf statischen Terminals. Eine erste Flexibilisierung des elektronischen Lernens brachte der Durchbruch des Personal Computers in den 1980er Jahren, der den Einsatz so genannter „Computer Based Trainings (CBT)“ auf Diskette oder CD-ROM ermöglichte. Die zweite, wesentliche Flexibilisierung des E-Learnings wurde durch den Einsatz des Internet erreicht. Dieses entwickelte sich seit Anfang der 1960er Jahre aus dem ARPANET, das vom amerikanischen Verteidigungsministerium voran getrieben wurde und als Forschungsnetzwerk diente. 1989 programmierte Tim Berners-Lee bei der Europäischen Organisation für die Kernforschung (CERN) in Genf den ersten Webserver und das erste Client-Programm und erfand damit das World Wide Web, einen der Dienste des Internet [JACOBS 2006]. Nachdem 1993 mit Mosaic der erste grafikfähige Browser vorgestellt wurde [CHANG 1993], wurde das World Wide Web auch bei Laien populär und die Nutzerzahlen nahmen exponentiell zu. Heute sind Übertragungsgeschwin-

digkeiten und Verbreitung des Internet so stark angewachsen, dass es zum primären Transportmedium für E-Learning geworden ist und die CD-ROM bzw. die CBTs abgelöst hat.

E-Learning wurde und wird in Deutschland von staatlicher Seite finanziell gefördert. Zu nennen sind hier beispielsweise das Computer-Investitionsprogramm, in dessen Rahmen seit 1985 so genannte „CIP-Pools“ entstanden und mit Rechnern ausgestattet wurden [SCHULMEISTER 2002]. Ein Förderprogramm neueren Datums ist das Projekt „Neue Medien in der Bildung“, das seit 2001 läuft und bis 2008 verlängert wurde [KINDT 2006].

2.1.3.2 Einfluss der Neuen Medien

Die so genannten Neuen Medien⁶ haben die für das E-Learning zur Verfügung stehenden technischen und didaktischen Möglichkeiten auf verschiedene Weisen verändert. Als Grundprinzip für den Einsatz dieser Medien gilt, dass ein Medium nicht zum Selbstzweck eingesetzt werden sollte, sondern dass durch die Nutzung des Mediums ein Mehrwert entsteht.

Bewegtmедien

Die einfachste Form eines im Rahmen von E-Learning einsetzbaren Bewegtmедiums ist das so genannte „*animated gif*“. Das Bildformat „gif“ („*graphics interchange format*“) erlaubt es, in einer Bilddatei statt eines Bildes mehrere Bilder abzulegen. Diese werden dann vom Anzeigeprogramm als Abfolge dargestellt. Dabei ist ein Anhalten oder Beschleunigen nicht möglich. Wenn sich die enthaltenen Bilder nur wenig voneinander unterscheiden, können dem Betrachter Zusammenhänge, die aus einer Abfolge von Ereignissen resultieren, vermittelt werden. Die beschriebene Technik ist heute weitgehend von hochwertigen Animationen überholt worden, hat aber, insbesondere aufgrund der einfachen Herstellung, noch ihre Daseinsberechtigung und kann didaktisch sinnvoll genutzt werden [WEHLING 2008] .

Mit speziellen Animationsprogrammen können Abläufe dargestellt werden, die zuvor nur mit viel Text und mehreren Bildern, die Zwischenstadien zeigten, vermittelt werden konnten. Animationen stellen somit eine didaktische Alternative bei der Erläuterung komplizierter Sachverhalte dar. Dies gilt auch für Simulationen, bei denen der Nutzer durch Parametervariation den Ablauf beeinflussen kann. Durch didaktische Vereinfachung können Zusammenhänge gegenüber Realfilmen besser herausgestellt und für das momentane Lehrziel unerwünschte Umstände (z.B. Ausnahmen von einer Regel) ausgeblendet werden. Jedoch ist die Erstellung von Animationen, insbesondere, wenn

6 Die „Neuen Medien“ werden in der Literatur im vorliegenden Zusammenhang auch als „Multimedien“ oder „Telemedien“ (vgl. z.B. [KERRES 2001]) bezeichnet. Gemeint sind in Anlehnung an WIKIPEDIA 2008 in jedem Falle digitale Medien und Dienste, die über das Internet zugänglich sind.

diese Interaktionen der Nutzer zulassen sollen, zeitlich aufwändig. Der De-Facto-Standard⁷ zur Erstellung von Animationen für die Veröffentlichung im Internet ist das Programm *Flash* der Firma *Adobe*.

Durch Videos können reale Abläufe, wie z.B. Versuchsdurchführungen, vermittelt werden, insbesondere, wenn ein tatsächlicher Praxisbezug nützlich ist. Auch hier sind durch Verkürzungen im Ablauf didaktische Vereinfachungen möglich bzw. sogar notwendig [WEHLING 2008]. Der Mehrwert von Realfilmen liegt vor allem darin, dass sie mehrfach verwendet werden können und der gefilmte Sachverhalt nicht mehrfach nachgestellt werden muss. Bei der Herstellung von Filmen müssen auch der Schnitt und das Exportieren in ein für das Internet geeignetes Format eingeplant werden. Einen Standard für das Format von Realfilmen gibt es nicht, es stehen also mehrere Möglichkeiten zur Verfügung.

Hypertexte

Der nicht-lineare, sondern durch Hyperlinks verzweigte Text wurde – wie oben beschrieben – schon vor dem Durchbruch des Internet erfunden. Er ist heute integraler Bestandteil des World Wide Web. Hypertexte ermöglichen theoretisch eine neue Art des Lesens, indem der Nutzer seinen Interessen durch Links zu anderen Texten folgen kann. Beim Abarbeiten von Lerneinheiten wird dabei die Entscheidung, auf welcher Seite man weiter lernen möchte, teilweise dem Lerner auferlegt. Ein Nachteil beim Einsatz von Hypertexten ist der so genannte „Lost-In-Hyperspace“-Effekt, im übertragenen Sinne kann sich ein Nutzer also im Internet oder in einem Lernsystem „verirren“. Konkret bedeutet das, dass der Lerner von seinem eigentlich intendierten Lernweg abweicht und dann entweder das Lernziel aus den Augen verliert oder den „Rückweg“ zur ursprünglichen Lerneinheit nicht findet [WEHLING 2008]. Umgekehrt betrachtet kann ein Abweichen von einem vorgegebenen Lernweg Studierenden auch neue, nicht vom Lehrenden geplante Erkenntnisse ermöglichen („Serendipitäts-Effekt“) [SCHULMEISTER 2002]

Bereitstellung und Publikation

Die Publikation von Dokumenten und multimedialen Inhalten über das Internet ist heute sehr einfach und auf unterschiedliche Weise möglich. Softwareprogramme zur Unterstützung von Web-Autoren stehen zur Verfügung. Der Aufwand für die Internetpublikation lässt sich mit dem für den herkömmlichen Skriptverkauf vergleichen. Lerninhalte, wie z.B. Skripte im PDF-Format, werden in der Regel auf der Homepage der jeweiligen Lehrinstitution bereitgestellt. Für umfangreiche Audio- oder Filmdateien gibt es die Möglichkeit, sie zum Download oder als so genanntes Streaming an-

⁷ Stand Juni 2008

zubieten. Letztere Möglichkeit hat den Vorteil, dass Dateien vor dem Beginn des Abspielens nicht komplett heruntergeladen werden müssen. Lernern kann z.B. durch Wiki-Systeme die Möglichkeit gegeben werden, eigene Meinungen oder Kommentare zu publizieren. Zu erwähnen ist in diesem Zusammenhang noch das so genannte „Web 2.0“, auf das im Folgenden noch näher eingegangen wird.

Da Informationen, auch fachlicher Natur, im Internet zur Genüge zu finden sind, stellt sich insbesondere für den fachlich unerfahrenen Nutzer das Problem eines Überangebots an Informationen. Wenn der Nutzer zusätzlich mit dem Medium Internet nicht vertraut ist und die Zuverlässigkeit aufgefundener Informationen nicht beurteilen kann [WHITMIRE 2004], verschärft sich diese Problematik. Im E-Learning wird durch Lernplattformen bzw. Learning-Management-Systeme versucht, die publizierten Informationen für den Nutzer übersichtlich zu verwalten und die Aktivitäten der Nutzer zwecks Lernkontrolle aufzuzeichnen. Lerner haben so stets den Überblick darüber, was sie bereits bearbeitet bzw. gelernt haben, und welche Inhalte noch bearbeitet werden müssen.

Bei im Rahmen von E-Learning publizierten Dokumenten muss zwischen authentischen und didaktisch aufbereiteten Dokumenten unterschieden werden [KERRES ET AL. 2002]. Die Publikation authentischer Dokumente wird durch die Neuen Medien zwar vereinfacht, stellt aber nach der Definition aus Abschnitt 2.1.1 noch nicht zwingend E-Learning dar.

Kommunikationsformen

Die Verbreitung des Internet brachte dessen Nutzern auch neue Möglichkeiten der Kommunikation. Allem voran hat die E-Mail die private und geschäftliche Kommunikation vereinfacht und beschleunigt. Dies ist im E-Learning bei der asynchronen Kommunikation wichtig, z.B. bei der Beantwortung von Fragen und beim Einsenden von Aufgaben. Weitere Formen der durch das Internet hervorgebrachten Kommunikation sind Foren und Chaträume. Chaträume erlauben den synchronen Austausch der Nutzer untereinander, sind also z.B. auch zum Abhalten von Sprechstunden geeignet. Da die Nutzung von Foren und Chats einfach und weitgehend anonym ist, halten sich Nutzer meist an Verhaltensregeln, die so genannte Netiquette.

Eine noch relativ junge Kommunikationsform im Internet ist das „Web 2.0“, umgangssprachlich auch als „Mitmachnetz“ bezeichnet. Es erlaubt den Nutzern, über webbasierte Formulare Inhalte auf fremden Webseiten zu publizieren. Dies kann von so genannten „Weblogs“ (Blogs) über Wiki-Webs zu verschiedenen Themen bis zum Lexikon „Wikipedia“ reichen. Gemein ist diesen Systemen, dass keine HTML-Inhalte extern erstellt werden müssen, sondern dass Text direkt in ein Eingabefeld auf einer Webseite eingegeben werden kann. Es sind also für den Autoren kein direkter Serverzugang und keine HTML-Kenntnisse erforderlich. Eine Vielzahl von Autoren kann so an der

Erstellung von Inhalten mitwirken. Im E-Learning sind solche Ansätze für den Austausch von Meinungen der Lerner untereinander oder für Projektarbeiten gut geeignet [HÄFELE U. MAIER-HÄFELE 2004]. Seiten nach dem Wiki-Prinzip sind auch als Autorenwerkzeug sinnvoll einsetzbar.

Didaktik

Mit der Anwendung der Neuen Medien gehen Änderungen der Didaktik einher. Die Neuen Medien ermöglichen durch die Selbststeuerung der Lerner und neue Kommunikationsformen ein „anderes Lernen“. Hierbei ist jedoch nicht nur das Material ausschlaggebend, das auf einer Plattform bereitsteht („google-findbar“), sondern vor allem die Strukturierung der Inhalte [KERRES ET AL. 2002]. Insbesondere selbstorganisiertes Lernen kann durch den Einsatz der Neuen Medien unterstützt werden [KERRES 2001].

Wenn Neue Medien im Rahmen selbstgesteuerten Lernens eingesetzt werden, ist die Motivation des Lerners besonders wichtig. Diese kann intrinsisch sein, also vom Lerner selbst ausgehen, z.B. durch Neugier oder Interesse an einem Thema. Oder sie wird von außen vorgegeben, z.B. durch den Zwang oder Wunsch, eine Prüfung zu bestehen. Eine solche Motivation wird als extrinsisch bezeichnet [GAGE 1986].

Sind die Lerninhalte eines E-Learning-Systems hinreichend vernetzt, können selbstorganisiert lernende Nutzer sich Inhalte unabhängig von äußeren Vorgaben individuell erschließen. Die Strukturierung der Lerninhalte kann durch eine elektronische Auswertung von Beziehungen zwischen verschiedenen Inhalten ermöglicht werden. Eine solche Vernetzung kann über ein Semantic Web auf Basis von Metadaten oder die Nutzung einer Taxonomie, die z.B. auf einer hierarchischen Struktur der Inhalte beruht, erzeugt werden [BERNERS-LEE ET AL. 2001]. Lerner, die sich in ein Themengebiet einarbeiten, können so mit ungenauen Anfragen an das System beginnen und sich durch die Vernetzung durch das Themengebiet bewegen. Eine genaue Kenntnis aller Fachbegriffe, nur um Anfragen an das System formulieren zu können, ist nicht erforderlich [RATH U. BAUMER 2003].

Auf Seiten der Lehrer bzw. Autoren müssen die strukturierenden Vorgaben wie Metadaten oder Taxonomien jedoch formalisiert und für die einzelnen Wissenseinheiten angegeben werden. An dieser Stelle setzt die vorliegende Arbeit an, indem gezeigt wird, wie auf der Basis lediglich der Inhalte sowie textstatistischer und linguistischer Informationen ebenfalls der Kontext von Inhalten in die Wissensvermittlung mit einbezogen werden kann.

2.1.3.3 Standardisierung

Im Bereich des E-Learning sind wegen des hohen Aufwands bei der Produktion von Inhalten Bemühungen zur Standardisierung im Gange. Verschiedene Organisationen und Firmen definieren Schnittstellen für eine übergreifende Nutzung von E-Learning-Inhalten. Wesentliche Aspekte dieser Standardisierung sind nach MONTANDON 2006

- Auffindbarkeit der Inhalte durch inhaltliche Beschreibung und Kategorisierung,
- Nachhaltigkeit, d.h. Wiederverwendbarkeit und Austauschbarkeit der Inhalte und
- Anpassbarkeit bzw. Erweiterbarkeit fremder Inhalte für die eigene Nutzung oder umgekehrt.

Es ist wichtig, dass die Standards die Produzenten von E-Learning-Inhalten nicht in ihren Gestaltungsmöglichkeiten, vor allem in didaktischer Hinsicht, einschränken [MONTANDON 2006]. Die zusammengekommenen Vorteile der Standardisierung bedeuten für den Produzenten oder Anwender von E-Learning natürlich auch eine höhere Wirtschaftlichkeit.

Als Lehrender ist man bei der Anwendung von standardisierten E-Learning-Inhalten relativ sicher, dass diese auch noch bei einer Erneuerung oder Änderung der verwendeten Lernplattform funktionieren. Dem Lernenden bietet sich der Vorteil, dass unterschiedliche E-Learning-Inhalte auf derselben Lernplattform angeboten werden können [SCHULMEISTER 2001] und die Einarbeitung in die Benutzung einer weiteren Plattform entfällt. Nach SCHULMEISTER 2003 kann die Nichtberücksichtigung von E-Learning-Standards als K.O.-Kriterium für den Einsatz einer Lernplattform an einer Hochschule gezählt werden.

Trotz der offensichtlichen Wichtigkeit der Standardisierung wird diese nicht von allen E-Learning-Anwendern genutzt. Jedoch beabsichtigt die Mehrheit der E-Learning-Projektverantwortlichen an deutschsprachigen Hochschulen, E-Learning-Standards zukünftig zu berücksichtigen (vgl. hier und im Folgenden [MONTANDON 2006]). Potentielle Nutzer von E-Learning-Standards sehen Plattformunabhängigkeit, Wiederverwendbarkeit und Nachhaltigkeit als wichtigste Kriterien, die für eine Berücksichtigung der Standardisierung sprechen. Dieselbe Studie offenbart, dass demgegenüber 62% der Befragten die Berücksichtigung der Standards als „komplex“ bezeichnen, 52% als „mühsam“ und 40% sogar als „frustrierend“.

Die bei E-Learning-Verantwortlichen bekanntesten Standards für E-Learning sind nach MONTANDON 2006 SCORM, die Dublin Core Metadata Initiative und IMS. Der derzeit am weitesten verbreitete Standard ist SCORM [ADL 2008]. Er beinhaltet Metadaten, die auf dem LOM-Standard („*Learning Object Metadata*“) des IEEE („*Institute of Electrical and Electronics Engineers*“) basieren. Dieser Metadaten-Standard gliedert sich in neun Kategorien. Er erfasst neben allgemeinen, technischen und pädagogischen Informationen zum beschriebenen Lernobjekt auch Beziehungen dieses Lernobjekts zu anderen Lernobjekten sowie dessen Einordnung in eine oder mehrere Kate-

gorien. Daneben ermöglicht SCORM auch die Bündelung und Sequenzierung von Lernobjekten. Als nachteilig ist anzusehen, dass der SCORM-Standard von vielen Anwendern und Entwicklern als zu restriktiv empfunden wird, da er u.a. das Einbinden von E-Learning-Elementen in andere Geschäftsprozesse von Institutionen bzw. Unternehmen erschwert [AROYO U. DICHEVA 2004].

2.1.3.4 Aktuelle Forschungsgebiete

Spätestens seit den Intelligenten Tutoring Systemen und auch heute noch ist die Adaptivität von E-Learning-Systemen an das Verhalten der Lerner ein verbreiteter Forschungsansatz. Dieser Ansatz folgt aus der Erkenntnis, dass die Diversität der Studierenden, vor allem in ihrem Lernverhalten, berücksichtigt werden muss. Jedoch kann tatsächliche Adaptivität an alle durch Lehrstoff und Lerner beeinflussten Variablen nach SCHULMEISTER 2004 nicht erreicht werden, da die Anzahl dieser zu berücksichtigenden Variablen zu hoch wäre. Ferner ist eine Anpassung der Lehrmethoden an die Lernmethoden wissenschaftlich noch nicht gesichert. Der Einsatz von Methoden der Künstlichen Intelligenz wird auf dem Gebiet der Adaptivität zunehmen. Die Neukombination von vorhandenen Lerneinheiten im Sinne sowohl der Nachhaltigkeit als auch der Adaptivität ist ebenfalls Ziel der Forschung.

Auch das Semantic Web, also die rechnerverständliche Vernetzung von natürlichsprachlichen Texten, wird als „Educational Semantic Web“ in die Entwicklung einbezogen [AROYO U. DICHEVA 2004]. Auf diese Weise soll Interoperabilität zwischen verschiedenen E-Learning-Systemen hergestellt werden können, was sowohl die Adaptivität der Systeme erhöht als auch der Mobilität der Studierenden nach der Vorgabe des Bologna-Prozesses zugute kommt. Weitere Forschungsspekte sind die Kommunikationsstandards von E-Learning-Systemen auf verteilten Servern.

Auf dem Gebiet der Didaktik geht es um die Einbeziehung des „Web 2.0“ in die Ausbildung von Studierenden. Auch Systeme, die den Lernern die Konstruktion eigenen Wissens ermöglichen, sind Ziel der didaktischen Forschung [SCHULMEISTER 2004].

2.1.4 Spezielle Formen des E-Learnings

In letzter Zeit sind vermehrt abgeleitete Konzepte des E-Learnings entstanden, deren Namen meist aus dem um ein englischsprachiges Adjektiv erweiterten Begriff „Learning“ bestehen. Die wichtigsten dieser Konzepte sind

- „Blended Learning“ (Wechsel von Präsenz- und E-Learning-Phasen, auch „hybride Lernarrangements“ nach KERRES ET AL. 2002),

- „Rapid Learning“ (didaktisch und technisch wenig anspruchsvolle Vermittlung sehr aktuellen Wissens),
- „Augmented Learning“ („angereichertes“ Lernen, Ergänzung der Präsenzlehre durch E-Learning [MÜHLHÄUSER 2004]) und
- „Ambient Learning“ („Lernen im Vorbeigehen“, auch „Ubiquitous Learning“: bezeichnet den unbeschränkten Zugang zu E-Learning Inhalten von jedem Ort mit verschiedenen Endgeräten [PARASKAKIS 2005]).

Das Konzept „Blended-Learning“ kann gerade im Grundstudium durch die Präsenzphasen einen nützlichen Zwang auf die Studierenden ausüben helfen, damit diese sich mit den E-Learning-Inhalten der betreffenden Lehrveranstaltung auseinander setzen [GRÜBL ET AL. 2003]. So kann z.B. die Bearbeitung bestimmter E-Learning-Inhalte als Voraussetzung für das Verständnis der nächsten Vorlesung festgelegt werden [WEHLING 2008].

2.1.5 Einsatz von E-Learning an Hochschulen

2.1.5.1 Allgemeiner Einsatz an Hochschulen

E-Learning wird heute in allen Bereichen des Bildungssektors angewendet. Dies beinhaltet die Schule bis zur Sekundarstufe, die Berufs- und Hochschulausbildung und die berufliche Weiterbildung. Neben den traditionellen Universitäten wird E-Learning auch an Fernuniversitäten eingesetzt. E-Learning in Schule und Berufsausbildung soll in der vorliegenden Arbeit nicht weiter untersucht werden. Betrachtet wird der Einsatz an der Hochschule und nach dem erfolgreichen Hochschulabschluss in der Weiterbildung.

Einen exponentiellen Zuwachs verzeichneten die E-Learning-Aktivitäten an deutschen Hochschulen mit dem BMBF-Förderprogramm „Neue Medien in der Bildung“. Von diesem Programm wurden alleine 541 Einzelprojekte in 100 Projektverbünden gefördert [KINDT 2006]. Hochschulübergreifende Projekte konnten sich auch ohne Förderung etablieren. Typische Beispiele für hochschulübergreifende E-Learning-Projekte sind Fernpraktika, die von mehreren Universitäten gemeinsam durchgeführt werden (vgl. z.B. [RÖHRIG U. JOCHHEIM 1999]).

Das Thema Weiterbildung ist seit einigen Jahren verstärkt in die gesellschaftliche Diskussion geraten. Neben privaten Unternehmen wenden sich auch Hochschulen dieser Thematik vermehrt zu. Lebenslanges Lernen wird heute jedem Arbeitnehmer angeraten, um mit der technischen Entwicklung mithalten zu können. Während eines „normalen“ Arbeitslebens muss zunehmend mehr ge-

lernt werden [KERRES ET AL. 2002]. Dies erfordert eine flexible Lernorganisation, für die E-Learning in vielen Fällen geeignet ist.

E-Learning kann sich an Hochschulen langfristig nur durchsetzen, wenn bei den beteiligten Hochschullehrern, die E-Learning einsetzen, auch eine entsprechende Lehrkompetenz vorhanden ist [KERRES ET AL. 2005] oder erreicht werden kann. Abbildung 2.1 zeigt die Nutzung verschiedener Varianten digitaler Medien in Lehrveranstaltungen. Man erkennt, dass „echtes“ E-Learning zwar noch nicht überall verbreitet ist, aber doch in einem großen Teil der Veranstaltungen zumindest mit Beamer und Notebook gearbeitet wird. Dies ist eine Grundlage, auf der der Einsatz von E-Learning ausgebaut werden kann.

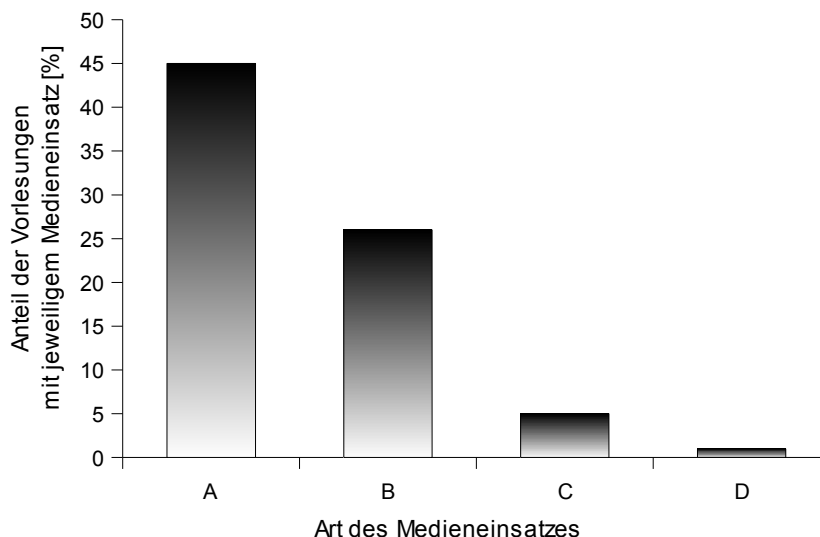


Abbildung 2.1: Anteil der Veranstaltungen, in denen verschiedene Varianten digitaler Medien eingesetzt werden (nach KERRES ET AL. 2005)

- A: Medieneinsatz in Präsenzveranstaltungen (Beamer, Notebook etc.)
- B: Medieneinsatz als Ergänzung zu normalen Präsenzveranstaltungen (Online-Materialien, Diskussions-Foren, virtuelle Sprechstunde etc.)
- C: Medieneinsatz als wesentliche Erweiterung zu Präsenzveranstaltungen (z.B. durch zeitlich alternierende Veranstaltungen, Selbstlernmodule)
- D: Medieneinsatz als Ersatz zu Präsenzveranstaltungen (z.B. Selbstlernprogramme, virtuelle Seminare)

Im Rahmen der Weiterbildung kommt es oft auf aktuelle Themen an, auch kurzfristig auftretender Bildungsbedarf muss bedient werden können [KERRES ET AL. 2002]. Man spricht hierbei auch von „Just-In-Time“-Wissensvermittlung [ABICHT U. DUBIEL 2003]. Dieses „Just-In-Time“-Lernen kann durch E-Learning oder auch Blended Learning besser als mit alleinigen Präsenzveranstaltungen

gen bedient werden [STOJANOVIC ET AL. 2001]. Dies liegt zum einen daran, dass in ein E-Learning-System jederzeit aktuelle Informationen aufgenommen und den Lernenden zugänglich gemacht werden können. Zum anderen sind E-Learning-Systeme schneller zugänglich, zur Vermittlung von Inhalten muss nicht der Termin einer Präsenzveranstaltung abgewartet werden.

2.1.5.2 E-Learning Projekte im Bauingenieurwesen

Die Vorteile des für viele Studiengänge sinnvoll einsetzbaren E-Learnings sind im Bauingenieurwesen als technischem Studiengang von besonderem Nutzen. In der Ausbildung von Ingenieuren haben Diagramme, Skizzen und Zeichnungen ein hohes Gewicht. Diese lassen sich gut multimedial darstellen. Weitere wesentliche Bestandteile in der Ausbildung der verschiedenen Fachdisziplinen des Bauingenieurwesens sind mathematische Verfahren [KÖHLER 2006]. Übungen, die Berechnungen zu verschiedenen Sachverhalten trainieren sollen, können gut im Rahmen von E-Learning abgebildet werden.

Sowohl durch Eigeninitiative als auch durch Projektförderung sind in den letzten Jahren an den Bauingenieur-Fachbereichen vieler deutscher Universitäten E-Learning Projekte entstanden. Eine Auswahl dieser Projekte inkl. einer kurzen inhaltlichen Beschreibung lässt sich Tabelle 2.3 entnehmen.

Auch die Bau- bzw. Baustoffindustrie stellt im Internet Animationen und Rechenprogramme zur Verfügung. Diese sollen natürlich in erster Linie die entsprechenden Produkte bekannt machen und ihren Verkauf steigern. Sie sind jedoch oft von hoher Qualität, fachlich korrekt und für die Ausbildung von Studierenden damit geeignet.

*Tabelle 2.3: E-Learning-Projekte aus dem Bauingenieurwesen an deutschen Hochschulen
(aus dem BMBF-Förderprogramm „Neue Medien in der Bildung“, Auswahl)*

Projektname	Fach / Fächer (Auswahl)	Hochschule(n)
Lernnetz Bauphysik	Bauphysik	Uni Karlsruhe* , TU Darmstadt, Uni Stuttgart, Uni Kassel, Uni Weimar, FH Biberach
WiBA-Net	Werkstoffe im Bauwesen	TU Darmstadt , TU Hamburg-Harburg, Uni Stuttgart, Uni Leipzig, Uni Duisburg-Essen, TU Berlin
FLUWU	Wasser und Umwelt	Uni Hannover , Uni Weimar, TU Braunschweig, TU Hamburg-Harburg, Uni München
Portiko	Baukonstruktion, Statik, Brandschutz, Konstruktiver Ingenieurbau, Baubetrieb	TU Braunschweig , TU Dresden
KI-SMILE	Konstruktiver Ingenieurbau	FH Potsdam , TFH Berlin, TU Berlin

* Die projektleitenden Hochschulen sind in fetter Schrift dargestellt.

2.1.6 Zusammenfassung: Aspekte des Einsatzes von E-Learning

E-Learning besitzt, verglichen mit herkömmlichen Lehrmethoden, i.d.R. Präsenzveranstaltungen, eine Reihe von Vor- und Nachteilen. Ein sinnvoller Einsatz von E-Learning in der Ausbildung von Studierenden oder in der Weiterbildung ist deshalb von individuellen Parametern abhängig.

Die in den vorausgegangenen Abschnitten herausgearbeiteten Vorteile von E-Learning sind:

- *Wiederverwendbarkeit von Inhalten:* Dies bedeutet, dass einmal erstellte Inhalte für unterschiedliche Fächer und an verschiedenen Hochschulen eingesetzt werden können. Insbesondere im Hinblick auf die Wiederholung von Grundlagen in Fächern höherer Semester ist dies auch ein interessanter didaktischer Aspekt.
- *Darstellung von Inhalten mit verschiedenen Medien:* Studierende haben unterschiedliche Lernverhalten. Durch die Variation der Medien kann einerseits versucht werden, einen möglichst großen Prozentsatz der Lerner auf ihrem bevorzugten Kommunikationsweg zu erreichen. Andererseits werden bei allen Studierenden unterschiedliche Aufnahmewege bedient, was zu einem besseren Behalten des Gelernten führen kann.
- *Verwendung von Bewegtmidien:* Bewegtmidien erlauben die Darstellung von Abläufen, die sonst umständlich als Text beschrieben und in statischen Bildern dargestellt werden müssen. Gut gemachte Animationen, Simulationen oder Filme können die Lerneffizienz und das Verständnis fördern.
- *Zeitersparnis bei Lehrer und Lerner:* Für Lerner entfallen ggf. Präsenzveranstaltungen sowie das Kopieren von Skripten. Versäumte Vorlesungen können leichter nachgeholt werden. Für den Lehrer ergibt sich die Zeitersparnis erst, wenn der E-Learning-Inhalt wiederholt eingesetzt werden kann. Die Zeit für die Erstellung der Inhalte ist jedoch bei einer Bilanzierung zu berücksichtigen.

Den beschriebenen Vorteilen können die folgenden Nachteile gegenüber gestellt werden:

- *Produktion erfordert hohen Aufwand:* Die Produktion von E-Learning-Inhalten, insbesondere von Animationen oder Filmen, erfordert einen hohen personellen, zeitlichen und finanziellen Aufwand. Ohne eine spezielle Förderung sind größere Projekte oft nicht umsetzbar.
- *Akzeptanzprobleme bei den Nutzern:* Sowohl Lehrende wie auch Lernende können von der Komplexität von E-Learning-Systemen abgeschreckt werden. Lernende können zwar zur Nutzung „gezwungen“ werden, jedoch sinkt bei schlecht gemachten oder als schlecht empfundenen E-Learning-Systemen die Motivation der Studierenden [SESINK 2003] und damit in der Regel auch der Lernerfolg [STEINBORN 2007]. Lehrende können z.B. durch die Komplexität von Autorenprogrammen von einer Nutzung abgehalten werden.

- *technische Probleme bei der Nutzung:* Bei der Nutzung von E-Learning-Inhalten, sei es als Lernender oder als Lehrender, können immer wieder technische Schwierigkeiten auftauchen. Diese können systemunabhängig sein, z.B. in Form einer fehlerhaften Internetverbindung, oder systembedingt, z.B. Schwierigkeiten beim Finden bestimmter Inhalte.

Bei der Abwägung der Vor- und Nachteile kann man die Effizienz von E-Learning als Verhältnis des Aufwands zum Ergebnis betrachten [KERRES ET AL. 2002]. Bei der Berücksichtigung ökonomischer Aspekte ist jedoch auch die Wiederverwendbarkeit, ggf. auch in der Weiterbildung, zu beachten.

Als wichtigstes Kriterium erfolgreicher Lernumgebungen gilt die Unterstützung der Individualität des Lerners [KERRES 1999] bzw. die Berücksichtigung der Diversität von Lernenden als Gruppe [SCHULMEISTER 2004]: Die Lernprozesse sollten im Idealfall auf Eigenaktivitäten des Lerners beruhen, was dem konstruktivistischen Modell der Didaktik entspricht. Die Lerner sollten so wenig wie möglich eingeschränkt werden und ihre persönlichen Lerninteressen sollten weitestgehend unterstützt werden.

2.2 Information-Retrieval

2.2.1 Grundlagen des Information-Retrieval

2.2.1.1 Aufgaben des Information-Retrieval

Das Information-Retrieval (IR) umfasst Bereiche der Informatik, des Bibliothekswesens, der Sprachwissenschaft und des Wissensmanagements. Es befasst sich mit der Erschließung von Informationen aus (elektronischen) Dokumentensammlungen. Die Erschließung beinhaltet Maßnahmen die Dokumentensammlung betreffend (z.B. Indizierung, Kategorisierung), die eigentliche Suche nach benötigten Informationen mittels verschiedener Verfahren sowie die Bereitstellung der Ergebnisse (z.B. Ordnen der Treffer, das so genannte „Ranking“). Unterschieden werden muss das Information-Retrieval von der Suche nach reinen Daten, z.B. in Faktendatenbanken [FERBER 2003] oder als kontextfreie Volltextsuche [BAEZA-YATES U. RIBEIRO-NETO 1999].

Anwendungsfelder des Information-Retrieval sind z.B. die Literaturrecherche in Bibliotheken, Recherchen in (Online-)Datenbanken, Suchmaschinen im Internet und nicht zuletzt das E-Learning, um Lernern und Lehrern zu ermöglichen, Informationen zu einem bestimmten Sachgebiet prägnant und strukturiert angezeigt zu bekommen bzw. diese Informationen in geeigneter Weise bereitstellen zu können. Für das entwickelte Programmsystem zur Lehrpfadgenerierung werden verschiedene Verfahren des modernen Information-Retrieval benutzt. Dies sind vor allem die automatische

Bestimmung von Beziehungen zwischen Inhalten aus dem gemeinsamen Auftreten von Begriffen, die darauf aufbauende Erweiterung von Suchanfragen um weitere passende Suchbegriffe sowie das Vektorraummodell.

Wenngleich die Wissenschaft des Information-Retrieval erst durch die Verbreitung des Internet und seiner Suchmaschinen breitere Aufmerksamkeit erfahren hat, so gibt es doch wesentliche Unterschiede zwischen dem so genannten „klassischen“ Information-Retrieval und dem Web Information-Retrieval. Die für die vorliegende Arbeit relevanten Unterschiede werden nachfolgend in Anlehnung an LEWANDOWSKI 2005 erläutert:

- Lernplattformen als Anwendungsfeld des klassischen IR, selbst wenn sie erweiterbar konzipiert sind, können zum Zeitpunkt einer konkreten Suchanfrage als „stationär“ betrachtet werden, das enthaltene Wissen wird vollständig erfasst. Eine vollständige Erfassung des im Internet enthaltenen Wissens ist demgegenüber heute technisch nicht möglich.
- Bei einer Internetrecherche besteht die Gefahr, redundante Antworten zu erhalten, beispielsweise durch Treffer, die dasselbe Dokument bezeichnen, die jedoch auf unterschiedlichen Servern bereitgestellt werden.
- Beim klassischen IR kann man davon ausgehen, dass die enthaltenen Inhalte, beispielsweise einer Lernplattform, fachlich korrekt sind und eine hinreichende Qualität aufweisen.
- Beim klassischen IR kann man von Nutzern ausgehen, die auch komplexere Suchmasken bedienen können und wollen. Internetsuchmaschinen haben meist ein sehr einfaches Benutzerinterface und verzichten deswegen auf einige Funktionen.
- Suchprogramme im klassischen IR können auf ein Themengebiet spezialisiert sein, Internetsuchmaschinen müssen für allgemeine Suchen geeignet sein.
- Die Rankingverfahren bei Internetsuche und klassischer Suche sind unterschiedlich, beispielsweise können bei der Suche nach Inhalten einer Lernplattform von menschlichen Experten erstellte Klassifizierungen von Dokumenten benutzt werden. Die Übergänge zu Webverzeichnissen (z.B. Yahoo⁸) sind allerdings fließend.

Obwohl das vorgeschlagene Lehrpfadgenerierungssystem über das Internet zu benutzen ist, kann man bei der Zusammenstellung der Ergebnisse von „klassischem“ Information-Retrieval ausgehen. Die Inhalte – im vorliegenden Fall die des Lernnetzwerks WiBA-Net – bestehen nur aus einer endlichen Menge von Dokumenten. Es kann davon ausgegangen werden, dass diese fachlich richtig sind, ihre Qualität ist also nicht in Frage zu stellen. Die einzelnen Dokumente enthalten nur in Ausnahmefällen Hyperlinks, die bei den Such- und Sortierv Verfahren zur Lehrpfaderstellung infolgedessen nicht berücksichtigt werden. Es sollen durch die implementierten Suchverfahren hauptsächlich

8 vgl. z.B. <http://de.dir.yahoo.com> (nachgewiesen am 27.05.2005)

Texte erschlossen werden. Die Inhaltsseiten, auf denen sich die Texte befinden, enthalten sehr oft auch Bilder oder Bewegtbildmedien. Für die Nutzung dieser Medien getrennt von den Texten steht eine Stichwortsuche auf Basis der Metadaten zur Verfügung.

2.2.1.2 Relevanz von Suchergebnissen

Das Kriterium zur Beurteilung der Qualität der auf eine Suchanfrage zurückgegebenen Dokumente ist deren Relevanz. Diese kann jedoch endgültig nur von menschlichen Experten beurteilt werden. Ziel bei der Entwicklung einer Suchstrategie muss es deswegen sein, dass die vom System auf eine bestimmte Suchanfrage hin zurückgegebenen Dokumente mit der Menge der von Experten als relevant eingestuften Dokumente möglichst genau übereinstimmen [BAEZA-YATES U. RIBEIRO-NETO 1999]. Die Relevanz eines Dokuments für eine bestimmte Suchanfrage hängt dabei nach FERBER 2003 ausschließlich von der Suchanfrage und dem Inhalt des Dokuments ab. Es ist unerheblich, ob bereits relevantere Dokumente mit demselben oder einem ähnlichen Inhalt gefunden wurden. Auch der Kenntnisstand des Suchenden wird bei der Relevanzbeurteilung nicht berücksichtigt. Ein weiterer Gesichtspunkt ist die korrekte Formulierung einer Suchanfrage. Geht man nicht von der Suchanfrage aus, zu der die zurückgelieferten Dokumente passen müssen, sondern vom tatsächlichen Informationsbedürfnis des Benutzers, das von den Ergebnissen befriedigt werden muss, so spricht man von „Pertinenz“ anstelle von „Relevanz“ [LEWANDOWSKI 2005].

Die Qualität einer Suchmaschine bzw. eines IR-Systems lässt sich im Wesentlichen durch die Relevanz der von ihr zurückgegebenen Suchergebnisse definieren. Jedoch ist eine objektive Bewertung der Relevanz wegen der nötigen Beurteilung durch menschliche Experten praktisch nicht möglich. Ein Lösungsansatz für dieses Problem ist die Durchführung von Suchanfragen auf speziell zusammengestellten Mengen von Dokumenten („Testkollektionen“). Die Leistung verschiedener Suchmaschinen kann dann durch einen Vergleich der jeweils gefundenen Dokumente quantifiziert werden. Solche Ansätze liefern jedoch kein absolutes Maß für die Leistung des Systems oder die Relevanz der Suchergebnisse [FERBER 2003]. Allerdings sind Maßzahlen für die Beurteilung der Relevanz definiert, die in der Praxis verwendet und im folgenden Abschnitt erläutert werden.

2.2.1.3 Precision und Recall

Als Evaluierungsmaße zur Beurteilung der Qualität eines Suchprogramms existieren die Merkmale „Precision“ und „Recall“. Precision („Genauigkeit“) bezeichnet den Anteil relevanter Dokumente an den vom Suchprogramm zurückgegebenen Dokumenten, Recall („Vollständigkeit“) den Anteil der

gefundenen relevanten Dokumente (vgl. hier und im Folgenden [FERBER 2003]). Die Bestimmung von Precision- und Recall-Wert ist in Abbildung 2.2 dargestellt. Idealerweise sind die beiden Werte natürlich gleich 1. Dieser Wert wird sich in der Praxis jedoch nur schwer erreichen lassen. Die beiden Größen verhalten sich unterschiedlich: Wird durch eine spezifischere Anfrage die Antwortmenge verkleinert, wird die Genauigkeit höher, die Vollständigkeit wird jedoch offensichtlich geringer.

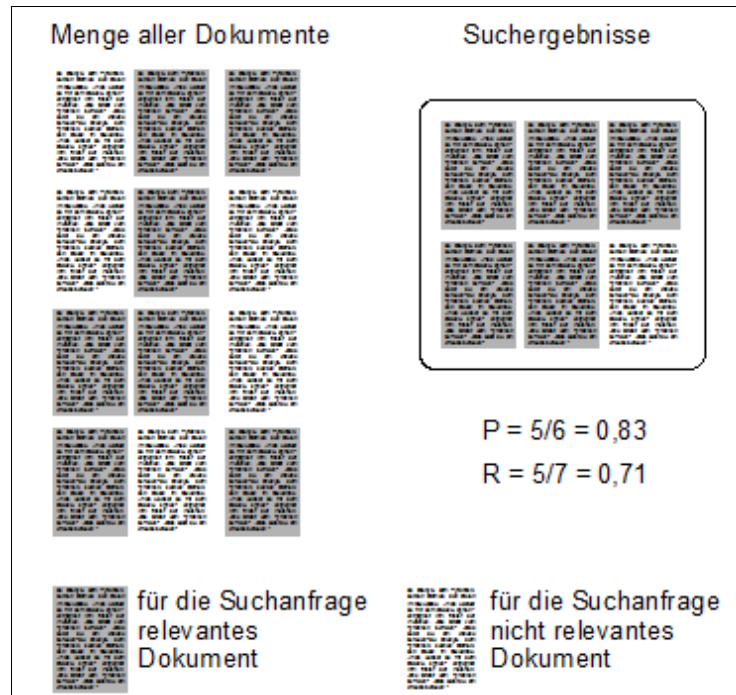


Abbildung 2.2: Precision und Recall (schematische Darstellung und Beispiel)

Precision und Recall einer Suchmaschine können, wie oben beschrieben, anhand von Testkollektionen von Dokumenten, für die exemplarische Suchanfragen und Informationen über relevante Dokumente existieren, überprüft werden. Die Werte einer Suchmaschine für Precision und Recall können beeinflusst werden, indem z.B. Grenzwerte modifiziert werden, ab denen zwei Dokumente oder ein Dokument und eine Suchanfrage als ähnlich betrachtet werden. Im Allgemeinen wird mit steigendem Recall die Precision kleiner und umgekehrt. Dies erschwert die Vergleichbarkeit von Suchmaschinen noch weiter.

2.2.2 Verfahren des Information-Retrieval

2.2.2.1 Indizierung und Boolesches Retrieval

Grundlage aller Verfahren des Information-Retrieval ist eine Indizierung der für die Suchanfragen zur Verfügung stehenden Textdokumente, d.h. die Speicherung der Informationen, in welchen Dokumenten und an welcher Stelle bestimmte Terme vorkommen. Am weitesten verbreitet ist die Indizierung mit Hilfe invertierter Listen. Die Invertierung besteht darin, dass nicht für ein Dokument festgehalten wird, welche Terme darin vorkommen, sondern dass jedem Term zugeordnet wird, in welchen Dokumenten er enthalten ist. Es wird also eine Liste aller Terme mit Verweisen auf die betreffenden Dokumente erstellt. Auf diese Weise verkürzt sich die Zeit zum Finden eines Dokuments.

In sehr großen Dokumentensammlungen können auch die invertierten Listen sehr umfangreich werden. Sie belegen dann viel Speicher und für die Suche wird viel Zeit benötigt, da zum Auffinden eines Terms große Datenmengen durchsucht werden müssen. Man behilft sich hier, indem die Informationen in zwei Dateien aufgeteilt werden. In der ersten Datei wird für jeden Suchbegriff gespeichert, an welcher Stelle in der zweiten Datei die benötigten Informationen stehen. Diese Datei braucht jetzt nur noch an dieser Stelle ausgelesen zu werden (vgl. hier und im Folgenden [FERBER 2003]). Bei einer Ergänzung oder Änderung des Datenbestands müssen die invertierten Listen neu kompiliert werden, was zusätzliche Rechenzeit erfordert. Dies ist bei einem erweiterbar konzipierten System zu berücksichtigen. Die invertierten Dateien können kleiner gehalten werden, wenn statt aller vorkommenden Wortformen nur deren Stammformen⁹ festgehalten werden, da in diesem Falle weniger Stichworte gespeichert werden müssen.

Das Boolesche Retrieval ist das vom Verfahren her einfachste Suchverfahren im Information-Retrieval. Überprüft wird, ob ein eingegebener Suchbegriff in genau derselben Form („exact match“) in den Dokumenten des Textkorpus vorkommt. Falls ja, wird das entsprechende Dokument als Suchergebnis zurückgegeben. Suchbegriffe lassen sich dabei mit den sog. Booleschen Operatoren „AND“, „OR“, „NOT“ und „XOR“ kombinieren.

Wird ein reines Boolesches Retrieval durchgeführt, so ist die Eingabe für ungeübte Nutzer schwierig [LEWANDOWSKI 2005]. Im Idealfall sollte der Katalog der möglichen Suchworte dem Suchenden bekannt sein, da ansonsten leere Mengen gebildet werden, wenn ein Suchbegriff gar nicht im System vorhanden ist. Ist der nicht vorhandene Suchbegriff durch einen AND-Operator mit einem anderen Begriff verknüpft, liefert die Suche ggf. gar keine Treffermenge. Vorteile des Booleschen Retrieval sind gute Ergebnisse mit hoher Relevanz, wenn die Suchanfrage korrekt formuliert wur-

9 vgl. Abschnitt 2.3.1

de, sowie die hohe Geschwindigkeit, die durch die Anwendung der invertierten Listen erreicht wird [FERBER 2003].

Beim Booleschen Retrieval wird lediglich untersucht, ob die Terme der Suchanfrage enthalten sind oder nicht. Es werden nur Dokumente zurückgegeben, die exakt auf diese Anfrage passen. Dies wird erschwert durch die Tatsache, dass die Suchbegriffe irgendwo im Text des gefundenen Dokuments stehen können und keinen inhaltlichen Bezug zueinander aufweisen müssen. Dokumente, die nur teilweise auf eine Suchanfrage passen, werden nicht zurückgegeben. Deswegen ist ein Ranking, also die Bestimmung einer Reihenfolge der Suchergebnisse, bei dieser Art des Retrievals nicht möglich. Jedoch ist es zumindest theoretisch möglich, die Suchbegriffe, durch logische Operatoren verknüpft, so exakt einzugeben, dass die Ergebnismenge hinreichend überschaubar wird. Es ist dann kein Ranking mehr erforderlich, da alle gefundenen Dokumente gesichtet werden können [LEWANDOWSKI 2005].

Die Übergänge vom Booleschen Retrieval zu anderen Methoden des Information-Retrieval sind fließend. So ist es z.B. möglich, auch abgeleitete Wortformen zu den Termen einer Suchanfrage aufzunehmen und damit die Menge der gefundenen relevanten Dokumente zu erhöhen.

2.2.2.2 Vektorraummodell

Das Vektorraummodell beschreibt ein Verfahren zur Bestimmung der Ähnlichkeit von zwei Dokumenten, oder auch zwischen einem Dokument und einer Suchanfrage. Das Modell geht davon aus, dass Dokumente als Vektoren in einem mehrdimensionalen Raum dargestellt werden können. Dabei ergibt sich die Anzahl der Dimensionen dieses Vektorraums aus den in den Dokumenten enthaltenen Termen. Jeder unabhängige Term stellt eine Dimension dar [LEWANDOWSKI 2005]. Es gibt verschiedene Methoden, die Ähnlichkeit zwischen diesen Vektoren zu bestimmen. Eine der einfachsten ist z.B. der Winkel zwischen ihnen: Je kleiner der Winkel wird, desto ähnlicher sind sich die Vektoren bzw. die Dokumente.

Das Vektorraummodell geht über das Boolesche Retrieval hinaus, denn hiermit lassen sich nicht nur Dokumente erfassen, die möglichst alle Terme einer Suchanfrage erfassen, sondern es kann auch die Ähnlichkeit der Dokumente mit der Anfrage berücksichtigt werden, sodass diese in eine bestimmte Reihenfolge gebracht werden können [FERBER 2003].

Zu den einer Inhaltsbasis angehörenden Dokumenten lassen sich Dokumentvektoren (D_i) erstellen. Diese sind in der Regel hochdimensional, jedoch schwach besetzt. Die Dimension entspricht dabei der Anzahl der relevanten Terme¹⁰ im Gesamtsystem. Vereinfachend kann auch die Vereini-

¹⁰ Die relevanten Terme sind die nach den in Abschnitt 4.2 beschriebenen Operationen verbleibenden sinntragenden Worte. Eliminiert werden z.B. Stoppworte wie „der“, „die“, „und“ etc.

gungsmenge der in zwei zu vergleichenden Dokumenten vorhandenen relevanten Terme betrachtet werden, hier sind die Dokumentvektoren jedoch für jede Vergleichsoperation neu zu bestimmen. Die einzelnen Stellen des Vektors bezeichnen das Auftreten eines bestimmten Terms im zugehörigen Dokument.

Zur Veranschaulichung sei hier der Aufbau eines Dokumentvektors anhand zweier aus je einem Satz bestehender Dokumente erläutert:

Satz 1: „Beton ist ein mineralischer Werkstoff.“

Satz 2: „Der Werkstoff Aluminium ist ein Metall.“

Die Vereinigungsmenge der beiden Sätze enthält insgesamt acht verschiedene Terme. Wenn man die beiden Dokumente im Rahmen des Vektorraummodells vergleicht, müssen also zwei Dokumentvektoren mit acht Dimensionen erstellt werden. Ordnet man die enthaltenen unterschiedlichen Begriffe alphabetisch, so ergibt sich für die Vereinigungsmenge „Aluminium“, „Beton“, „Der“, „ein“, „ist“, „Metall“, „mineralischer“, „Werkstoff“. Dieser Zeilenvektor lässt sich transponiert als Vektor schreiben:

$$\begin{pmatrix} \text{Aluminium} \\ \text{Beton} \\ \text{Der} \\ \text{ein} \\ \text{ist} \\ \text{Metall} \\ \text{mineralischer} \\ \text{Werkstoff} \end{pmatrix}$$

Es ergeben sich also die beiden Dokumentvektoren

$$D_1 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 1 \\ 0 \\ 1 \\ 1 \end{pmatrix}, \quad D_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 1 \end{pmatrix}$$

Diese weisen an den Stellen für „ein“, „ist“ und „Werkstoff“ jeweils beide den Wert 1 auf.

Die Dokumentvektoren stellen eine mathematische Formulierung von Texten dar, und ermöglichen, diese objektiv anhand gemeinsam vorhandener Begriffe zu vergleichen. Für diesen Ver-

gleich gibt es verschiedene Methoden. Man berechnet entweder die Ähnlichkeit zwischen zwei Vektoren oder die Entfernung zwischen den beiden Endpunkten der Vektoren im n -dimensionalen Raum [LEE 1999]. Für unterschiedliche Einsatzzwecke können verschiedene Verfahren der Ähnlichkeits- bzw. Distanzbestimmung unterschiedlich gut geeignet sein.

Grundsätzlich entspricht die Dimension der Dokumentvektoren D_i der Menge aller im zugrunde liegenden Textkorpus enthaltenen relevanten Schlagworte n . Die relevanten Schlagworte bestehen aus allen Worten des Volltextes, abzüglich der am häufigsten vorkommenden Worte. Die häufigsten Worte treten in zu vielen Einzeltexten auf, um als Suchbegriff geeignet zu sein und werden deshalb als so genannte „Ausschluss-“ oder „Stoppworte“ nicht berücksichtigt. Werden neue Wissensseinheiten in die Inhaltsbasis aufgenommen, so wird der Textkorpus vergrößert. Die Menge der neuen relevanten Schlagworte, die vorher noch nicht im Textkorpus enthalten waren, sei m . Dann wird die Menge der Schlagworte um m neue Ausdrücke erhöht. Die Dimension aller Vektoren w verändert sich dann, jedoch sind die letzten $n - m$ Komponenten der Vektoren gleich null, da die entsprechenden Schlagworte nicht vorhanden sind. Der Betrag eines Vektors w ändert sich dadurch nicht: Er beruht auf der Quadratsumme der einzelnen Elemente des Vektors, die durch neu hinzukommende Elemente mit dem Wert null nicht verändert wird.

In der Praxis können je nach beabsichtigter Berechnung Dokumente unmittelbar vor der Ähnlichkeits- oder Distanzbestimmung in Vektoren umgewandelt werden. Die Dimension dieser Vektoren richtet sich nach der Vereinigungsmenge der enthaltenen relevanten Terme, sie ist für zwei zu vergleichende Vektoren jedoch immer gleich. Sie ist außerdem i.d.R. um ein Vielfaches geringer als die durch die Menge aller relevanten Schlagworte vorgegebene Dimension n . Da oft große Mengen von Dokumenten miteinander zu vergleichen sind, ergibt sich durch diese Reduzierung der Dimension eine Verkürzung der benötigten Rechenzeit.

2.2.2.3 Erweiterung von Suchanfragen

Ziel einer Suchfunktion ist die Rückgabe möglichst aller für das Informationsbedürfnis des Nutzers relevanten Dokumente. Jedoch umfassen Suchanfragen nicht immer ausreichend Terme, um ein befriedigendes Suchergebnis zurückliefern zu können. Aus diesem Grund wird im Information-Retrieval versucht, die vom Nutzer vorgegebenen Suchbegriffe sinnvoll so zu erweitern, dass ein größerer Anteil relevanter Dokumente gefunden wird. Es gibt zwei grundlegende Möglichkeiten der Suchterweiterung: die Erweiterung um weitere grammatische Formen des eingegebenen Begriffs bzw. um semantisch ähnliche Worte und die Erweiterung aus dem Kontext.

Die erste Form der Erweiterung beinhaltet zunächst abgeleitete Formen der Suchbegriffe. Da bei der Beugung von Worten im Deutschen häufig Umlautungen oder Vokalverschiebungen vorkommen und für viele Konjugations- oder Deklinationsfälle eigene Endungen an Worte angehängt werden, muss bei einer Booleschen Suche auch nach diesen Formen gesucht werden. Ferner können Synonyme der Suchbegriffe aus einer entsprechenden Datenbank entnommen und ebenfalls mit ihren entsprechenden Ableitungen der Suchanfrage hinzugefügt werden.

Eine darüber hinausgehende Erweiterung des Suchvektors anhand ähnlicher Themen („Erweiterung aus dem Kontext“) kann mit Hilfe der Signifikanz des gemeinsamen Auftretens von Worten („Kookkurenzen“, vgl. Abschnitt 2.3.2) erfolgen. Es ist auch möglich, häufig mit den aktuellen Suchbegriffen gemeinsam eingegebene Suchbegriffe zur Suchterweiterung zu nutzen. Bei manchen Suchsystemen wird der Nutzer auch gefragt, ob er der Erweiterung der Suchanfrage zustimmt.

Aus den mit einem der beschriebenen Verfahren gefundenen Termen wird dann ein erweiterter Suchvektor erstellt. Dessen Einträge werden mit der inversen Dokumentdatei abgeglichen. Die Übereinstimmungen liefern dann die Suchergebnisse als Namen der Dateien, die die Suchbegriffe enthalten.

2.2.2.4 Linguistische Verfahren

Um die Verfahren des Information-Retrieval vollständig aufzuführen, müssen auch die linguistischen oder textstatistischen Verfahren genannt werden. Sie beruhen auf der Auswertung der Häufigkeit des Vorkommens von Worten oder Wortpaarungen in den einzelnen Dokumenten einer Kollektion. Da ein Schwerpunkt der vorliegenden Arbeit die Anwendung linguistischer Verfahren ist, werden diese in Abschnitt 2.3.2 im Zusammenhang mit dem Thema Computerlinguistik detailliert vorgestellt.

2.2.2.5 Weitere Verfahren

Neben den bisher vorgestellten, im Rahmen dieser Arbeit angewendeten Verfahren gibt es noch weitere, die in der Praxis angewendet oder in der Theorie diskutiert werden.

Das sicherlich bekannteste Verfahren des Information-Retrieval ist das von der Suchmaschine Google angewandte Page-Rank-Verfahren [BRIN U. PAGE 1998]. Hierbei werden nicht allein die Inhalte der untersuchten Seiten in das Ranking einbezogen, sondern vor allem die auf diese Seite verweisenden Linkstrukturen [FISCHER 2006]. Das Verfahren ist so bedeutend, dass Webseiten daraufhin optimiert werden, um in der Trefferliste von Google möglichst weit oben zu landen.

Ein weiteres, wenn auch hauptsächlich in der Theorie des Information-Retrieval wichtiges Verfahren ist das probabilistische Verfahren. Hierbei wird versucht, die Wahrscheinlichkeit der Relevanz eines Dokuments für eine Suchanfrage zu bestimmen. Dies geschieht anhand der im Suchvektor und im Dokument vorkommenden Worte und deren Häufigkeiten [LEWANDOWSKI 2005].

2.2.3 Mathematische Grundlagen für das Information-Retrieval

2.2.3.1 Ähnlichkeitsmaße

Bei der Anwendung des Vektorraummodells kommt es darauf an, Ähnlichkeiten zwischen zwei Dokumenten bestimmen zu können. Dies geschieht entweder, um eine Reihenfolge auf der Basis größer werdender Ähnlichkeiten festlegen zu können („Ranking“) oder um die Zusammengehörigkeit zweier Dokumente quantifizieren zu können.

Die zu vergleichenden Dokumente können dabei zusammenhängende Texte sein oder aus einzelnen Worten bestehende Vektoren, im Sinne der vorliegenden Arbeit also Suchvektoren. Auch für zwei alleine stehende Begriffe lässt sich eine Ähnlichkeit berechnen. Dazu müssen die Begriffe durch eine Reihe von ähnlichen Worten charakterisiert werden, beispielsweise durch ihre Kookkurrenzen. Die Ähnlichkeit der Begriffe lässt sich dann anhand der aus diesen Wortlisten gebildeten Vektoren bestimmen.

Ein gängiges und zugleich einfach zu bestimmendes Ähnlichkeitsmaß ist das Kosinusmaß. Hierbei wird der Kosinus des Winkels zwischen zwei Dokumentenvektoren bestimmt. Je höher dieser Wert ist, desto größer ist die Ähnlichkeit zwischen den betrachteten Dokumenten. Für den Kosinus-Wert und damit die Ähnlichkeit zweier Vektoren gilt:

$$\cos(\vec{a}, \vec{b}) = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| \cdot |\vec{b}|} \quad (2.1)$$

Für das oben begonnene Beispiel mit den Dokumentenvektoren D_1 und D_2 kann mit dieser Gleichung eine als Zahl ausdrückbare Ähnlichkeit bestimmt werden. Es gilt dann:

$$\cos(D_1, D_2) = \frac{D_1 \cdot D_2}{|D_1| \cdot |D_2|} \quad (2.2)$$

$$\cos(D_1, D_2) = \frac{3}{\sqrt{5} \cdot \sqrt{6}} = 0,547 \quad (2.3)$$

Nicht alle Terme können ein Dokument gleich gut beschreiben. Um genauere Werte für die Ähnlichkeit bestimmen zu können, können die einzelnen Terme im Vektor noch gewichtet werden. Dieses Vorgehen wird detailliert in Abschnitt 2.3.2 erläutert.

Wenn man nicht die Ähnlichkeit zwischen zwei Dokumenten, sondern ihre Entfernung, die „euklidische Distanz“, bestimmt, spricht man von einem „Distanzmaß“. Bei einem Distanzmaß gilt, dass die Ähnlichkeit mit kleiner werdender Distanz größer wird. Weitere Maße für die Ähnlichkeit sind z.B. das Skalarprodukt der Dokumentvektoren, das Dice-Maß oder das Jaccard-Maß, die sehr ähnlich berechnet werden, oder das Overlap-Maß [FERBER 2003]. Bei manchen Dokumentenvergleichen müssen sehr viele Dokumente miteinander verglichen werden. Deshalb muss bei der Bestimmung der Eignung eines Ähnlichkeitsmaßes für eine bestimmte Aufgabe auch die Anzahl der durchzuführenden Rechenoperationen und damit die resultierende Rechenzeit berücksichtigt werden.

Im Folgenden wird für die Ähnlichkeit zwischen zwei Dokumentvektoren D_i und D_j die Schreibweise $\text{sim}_{\text{di}}(d_j)$ (von engl. „similarity“) verwendet.

2.2.3.2 Clustering

Clustering bezeichnet die Zusammenfassung verschiedener Dokumente einer Dokumentenmenge (z.B. Suchergebnisse) zu inhaltlich ähnlichen Gruppen. Die Gruppen ergeben sich erst direkt beim Clustering. Dies unterscheidet das Clustering von der Klassifizierung, bei der die Kategorien, denen die Dokumente zugeordnet werden, schon vorab festgelegt sind [LEWANDOWSKI 2005].

Für das Clustering von Dokumenten gibt es verschiedene Verfahren. Grundsätzlich können dabei in Anlehnung an FERBER 2003 zwei Ansätze unterschieden werden. Zum einen das Erstellen eines Clusters aus jedem Dokument und die nachfolgende Zusammenfassung der zwei jeweils ähnlichsten Cluster. Zum anderen das Festlegen eines Dokuments als erstem Clusterkern, dem dann solange die verbleibenden Dokumente, sortiert nach deren Ähnlichkeit zum Cluster, zugefügt werden, bis die Ähnlichkeit eines Dokuments zum Cluster zu gering wird. Erst dann wird der nächste Cluster angelegt (vgl. Abbildung 2.3).

Auch für das Clustering können grundsätzlich verschiedene Ähnlichkeitsmaße angewendet werden. Es können zum einen die Schwerpunktsabstände zwischen Cluster und Dokument bestimmt werden, zum anderen auch die Ähnlichkeiten zwischen dem Schwerpunktsvektor eines Clusters und einem Dokumentvektor. Bei der Verwendung eines Entfernungsmaßes wird eine Seite einem Cluster hinzugefügt, wenn ein festgelegter Grenzwert unterschritten wird. Wird ein Maß, das auf

dem Winkel zwischen den betrachteten Vektoren basiert, verwendet, so wird die Seite ins Cluster aufgenommen, wenn der festgelegte Grenzwert überschritten wird.

Der Schwerpunkt eines einzelnen Dokuments entspricht dessen Dokumentvektor. Der Schwerpunkt eines Clusters berechnet sich aus den Dokumentvektoren der im Cluster enthaltenen Dokumente. Dazu werden die Dokumentvektoren addiert und der sich ergebende Vektor mit dem Kehrwert der Anzahl der Dokumentvektoren multipliziert. Nach jedem Zufügen eines Dokuments zu einem Cluster muss dessen Schwerpunkt neu berechnet werden.

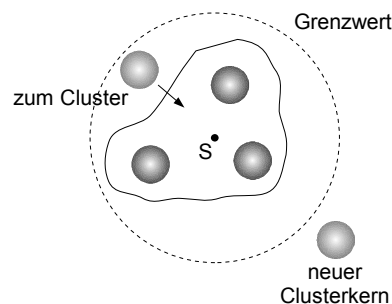


Abbildung 2.3: Zuordnung von Dokumenten zu einem bestehenden Cluster mit dem Schwerpunkt S auf der Basis der Entfernung der Dokumente zum Cluster

2.2.3.3 Fuzzy Logic

Durch Fuzzy Logic wird die mathematische Mengenlehre von exakt definierten auf unscharf (= „fuzzy“) umrissene Mengen erweitert. In der herkömmlichen Mengenlehre gehört ein Element entweder zu einer Menge oder es gehört nicht dazu, Zwischenzustände sind nicht möglich. In der unscharfen Mengenlehre kann ein Element auch zu einem bestimmten Zugehörigkeitsgrad einer Menge angehören. Die „normale“ Mengenlehre lässt sich als Spezialfall der unscharfen Mengenlehre definieren [FERBER 2003]. Auf diese Weise kann die Ausdrucksweise der menschlichen Sprache in die Datenverarbeitung einfließen, die auch Empfindungen und Einschätzungen beinhaltet [TRAEGER 1994]. Mit Hilfe der Fuzzy Logic lassen sich unscharfe Begriffe modellieren, die durch den „gesunden Menschenverstand“ einfach zu erfassen, computerverständlich aber nur schwer darstellbar sind. Die so modellierten Begriffe lassen sich dann durch Fuzzy-Systeme berechnen [KARAGIANNIS U. TELESKO 2001].

Unscharfe Mengen werden nach linguistischen Variablen benannt, nämlich Begriffen der Umgangssprache wie „hoch“, „niedrig“ oder auch „thematisch zur Suchanfrage passend“. Die Zugehörigkeit eines Elements x zu einer unscharfen Menge A wird durch den Zugehörigkeitsgrad $\mu_A(x)$ ausgedrückt (vgl. Abbildung 2.4).

Im Folgenden soll anhand von Rohdichte und Druckfestigkeit der Baustoffe Stahl und Beton die Vorgehensweise bei einer Fuzzy-Berechnung erläutert werden. Für die Rohdichten der beiden Baustoffe gilt:

$$\rho_{\text{Stahl}} = 7.850 \text{ kg/m}^3$$

$$\rho_{\text{Beton}} = 2.400 \text{ kg/m}^3$$

Zugehörigkeitsgrade liegen im Intervall $[0,1]$. Hinsichtlich der Rohdichte von Baustoffen könnte beispielsweise festgelegt worden sein¹¹:

$$\mu_{\text{schwer}}(\text{Stahl}) = 0,9$$

$$\mu_{\text{schwer}}(\text{Beton}) = 0,3$$

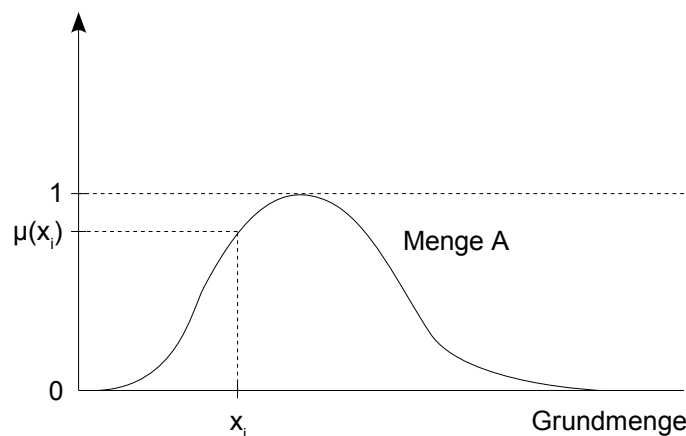


Abbildung 2.4: Grafische Darstellung des Zugehörigkeitsgrades μ des Elements x_i zur Menge A (nach TRAEGER 1994)

Es gelten folgende Rechenregeln für Multiplikation und Addition von Zugehörigkeitsgraden zu unscharfen Mengen:

Multiplikation:

$$\mu_{A \cdot B}(x) = \mu_A(x) \cdot \mu_B(x) \quad (2.4)$$

Addition:

$$\mu_{A+B}(x) = \mu_A(x) + \mu_B(x) - \mu_A(x) \cdot \mu_B(x) \quad (2.5)$$

Zur Durchführung von unscharfen Berechnungen muss eine vorliegende Problemstellung zunächst in Fuzzy-Mengen übertragen werden. Die scharfen, reellwertigen Eingangs- und Ausgangswerte werden dabei in linguistische Variablen übersetzt [KARAGIANNIS U. TELESKO 2001]. So könnte

¹¹ Es handelt es sich hier um Schätzwerte zur Veranschaulichung. Weiter unten werden diese Zugehörigkeitsgrade auf der Basis von Zugehörigkeitsfunktionen berechnet.

z.B. der scharfen Eingangsgröße „20 g/dm³“ die linguistische Variable „leicht“ zugeordnet werden. Dieser Vorgang wird als „Fuzzyfizierung“ bezeichnet. Zur Fuzzyfizierung eines Wertes wird eine Zugehörigkeitsfunktion aufgestellt, die reellen Zahlenwerten Zugehörigkeitsgrade zuordnet. Sie kann einen beliebigen Verlauf haben, muss den zu übersetzenden Sachverhalt aber sinnvoll abbilden [TRAEGER 1994]. Die Zugehörigkeitsgrade für die Eingangswerte lassen sich dann an der Zugehörigkeitsfunktion ablesen oder mit dieser berechnen. Erfahrungsgemäß ist es praktisch, die Zugehörigkeitsfunktionen so zu wählen, dass die Zugehörigkeitsgrade wie im Beispiel die Summe 1 ergeben [TRAEGER 1994].

Als Fortführung des obigen Beispiels sei wieder die Rohdichte von Baustoffen betrachtet. Es kann hier beispielsweise unterschieden werden zwischen den Eigenschaften „leicht“, „mittel“ und „schwer“ (vgl. Abbildung 2.5).

Für die o.a. Rohdichten zweier Baustoffe, angegeben als konkrete Zahlenwerte, kann nun eine Zuordnung zu den unscharfen Mengen vorgenommen werden. Im vorliegenden Beispiel hätte Stahl mit einer Rohdichte von 7.850 kg/m³ also einen Zugehörigkeitsgrad von 1,0 zur unscharfen Menge „schwer“ ($\mu_{\text{schwer}}(7850) = 1,0$). Beton gehört mit einer Rohdichte von 2.400 kg/m³ zu 0,53 zur unscharfen Menge „leicht“ und zu 0,47 zur unscharfen Menge „mittel“. Analog können jetzt die Druckfestigkeiten der betrachteten Baustoffe den unscharfen Mengen „geringe Druckfestigkeit“, „mittlere Druckfestigkeit“ und „hohe Druckfestigkeit“ zugeordnet werden. Für das Beispiel habe der Baustoff Beton eine Zugehörigkeit von 0,7 zur Menge „hohe Druckfestigkeit“.

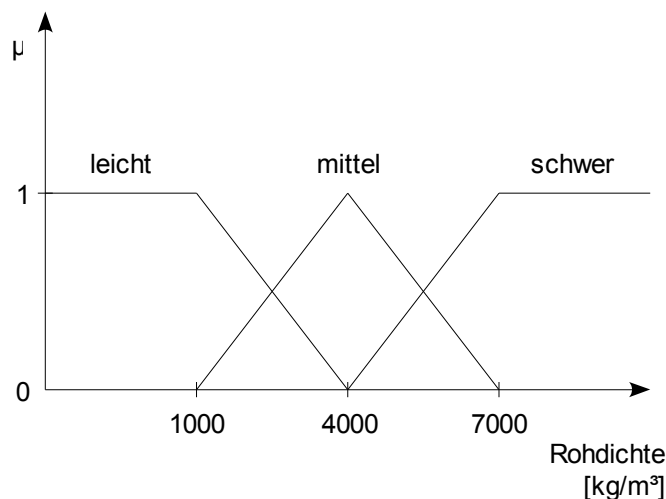


Abbildung 2.5: Fuzzy-Mengen für Rohdichte (Beispiel)

Die eigentliche unscharfe Berechnung wird „Inferenz“ genannt. Hierbei werden verschiedene Inferenzregeln, so genannte „Verarbeitungsregeln“, die sich auf die unscharfen Eingangsgrößen beziehen, nacheinander abgearbeitet (vgl. hier und im Folgenden [TRAEGER 1994]). Die Inferenzregeln

bestehen aus Prämissen, die untereinander verknüpft sein können, und Schlussfolgerungen, die eintreten, wenn die jeweilige Prämisse wahr ist. Durch die Anwendung der Regeln werden die Zugehörigkeitsgrade der Ergebnisteilmengen bestimmt. Eine Inferenzregel ist nach folgendem Muster aufgebaut:

WENN Voraussetzung₁ (UND/ODER Voraussetzung₂) DANN (Schlussfolgerung).

Als Fortsetzung des obigen Beispiels soll eine – stark vereinfachte – Einsatzempfehlung für die Baustoffe in Abhängigkeit ihrer Rohdichte und ihrer Druckfestigkeit gemacht werden. Zur Verfügung stehen dabei die Kategorien „nicht empfehlenswert“, „empfehlenswert“ und „sehr empfehlenswert“. Eine der Inferenzregeln für die Rohdichte eines Baustoffs und dessen Druckfestigkeit könnte dann lauten:

WENN (Rohdichte = „leicht“) UND (Druckfestigkeit = „hoch“) DANN (Einsatz = „sehr empfehlenswert“).

Durchschnitt und Vereinigung von Mengen bzw. die Verknüpfungen UND und ODER werden in der Praxis durch den Minimum-Operator respektive den Maximum-Operator abgebildet [KARAGIANNIS U. TELESKO 2001]. Die als Beispiel angegebene Inferenzregel wird also zu $\mu_{\text{sehr empfehlenswert}}(\text{Beton}) = \min(0,53; 0,7) = 0,53$. Da derselbe Baustoff auch noch zu gewissen Graden den Mengen „mittlere Rohdichte“ bzw. „mittlere Druckfestigkeit“ angehört, können unter Anwendung entsprechender Inferenzregeln auch für die Mengen „empfehlenswert“ bzw. „nicht empfehlenswert“ Zugehörigkeiten bestimmt werden.

Es gibt verschiedene Methoden, um aus der Abarbeitung der Regeln die Zugehörigkeitsgrade der Ergebnisteilmengen zu bestimmen. Eine der gebräuchlichsten ist die so genannte „Methode der Teilschwerpunkte“ [TRAEGER 1994]. Hierbei werden die unscharfen Ergebnismengen gleichsam als „Dicke“ des jeweiligen Bereichs der Zugehörigkeitsfunktion der Ergebnisteilmenge betrachtet. Die Ergebnismengen für den Baustoff Beton aus dem Beispiel sind in Abbildung 2.6 schraffiert dargestellt.

Der letzte Schritt bei der Anwendung der Fuzzy Logic ist die Defuzzifizierung. Hierbei werden die unscharfen Ergebnismengen wieder in einen scharfen Ausgabewert umgerechnet. Für die Defuzzifizierung gibt es verschiedene Methoden. Die angewandte Methode sollte der bei der Fuzzy-Inferenz zur Bildung der Ergebnisteilmengen verwendeten entsprechen. Bei der Methode der Teilschwerpunkte werden die Schwerpunkte aller Teilflächen berechnet (also nicht nur des schraffierten Bereichs). Die jeweilige Zugehörigkeit wird bei der Berechnung des Gesamtschwerpunkts dann als „Dicke“ der Fläche angesetzt [TRAEGER 1994].

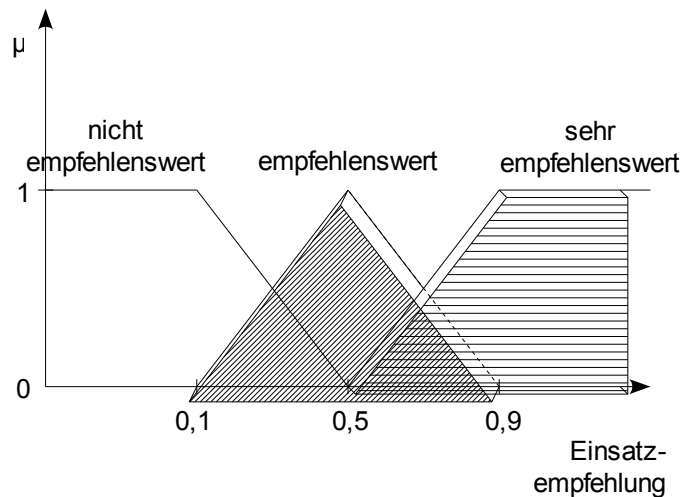


Abbildung 2.6: Fuzzy-Mengen für die Einsatzempfehlung

Im Beispiel kann für die beiden Baustoffe Stahl und Beton jetzt jeweils der Gesamtschwerpunkt ihrer Ergebnisteilmengen berechnet werden. Dieser hängt wegen der bereits feststehenden Schwerpunkte der Teilflächen nur von der Dicke der Teilflächen ab. Der x-Wert des Schwerpunktes stellt das Ergebnis der Fuzzy-Berechnung dar. Der Baustoff, dessen Einsatzempfehlung den höchsten scharfen Zahlenwert aufweist, sollte folglich eingesetzt werden. Das Beispiel wurde zur Erläuterung der Vorgehensweise bei einer unscharfen Berechnung natürlich stark vereinfacht.

2.3 Sprachwissenschaft

2.3.1 Linguistische Grundlagen

An dieser Stelle soll eine kurze Einführung in die Aspekte der deutschen Sprache, die für die vorliegende Arbeit relevant sind, gegeben werden. Dies betrifft insbesondere die schriftlich kodierte Form der Sprache.

Die Elemente der Sprache, die im Rahmen dieser Arbeit untersucht werden, sind in erster Linie Worte. Deshalb soll zunächst eine Definition des Begriffs Wort gegeben werden. Betrachtet man eine Buchstabenfolge, die von Leerzeichen umgeben ist, so spricht man von einem „orthographischen Wort“ (vgl. hier und im Folgenden [MEIBAUER ET AL. 2002]). In der gesprochenen Sprache tritt noch das phonologische Wort hinzu, das sich durch Sprechpausen definieren lässt. Eine für die vorliegende Arbeit relevante Definition ist vor allem das „morphologische Wort“. Dieses kann entweder aus einem freien Morphem bestehen und wird dann als Simplizium bezeichnet, oder aus einer frei auftretenden Morphemkonstruktion, man spricht dann von einem komplexen Wort.

Morpheme stellen die minimale bedeutungstragende Einheit der Sprache dar [TROMMER 2001]. Man unterscheidet freie und gebundene Morpheme. Freie Morpheme können für sich stehen, gebundene nicht. Zu letzteren zählen u.a. Flexionsendungen, wie z.B. „-e“ als Pluralmarkierung. Unterschiedliche Wortformen eines Wortes, z.B. mit einem unterschiedlichen Kasus, werden auch als syntaktische Wörter bezeichnet [MEIBAUER ET AL. 2002]. Aus Grundmorphemen, auch „Wurzeln“ oder „Stämme“ genannt, können durch „Flexion“ gebeugte Formen und durch „Derivation“ Ableitungen gebildet werden. Dies geschieht durch Hinzufügen eines „Affixes“, das der Wurzel entweder voran gestellt („Präfix“) oder angehängt („Suffix“) wird [KÖNIG 2001]. Eine Übersicht der Affixe im Deutschen findet sich in MEIBAUER ET AL. 2002. Bei der Derivation eines Wortes durch Hinzufügen eines Präfixes („Präfigierung“) tritt häufig eine Bedeutungsänderung des ursprünglichen Wortes auf, z.B. bei der Verwendung der Präfixe „un-“ oder „ent-“. Die Änderung der Bedeutung kann ohne einen entsprechenden Thesaurus weder erkannt noch spezifiziert werden. Im Rahmen der vorliegenden Arbeit werden also ausschließlich flektierte Wortformen und durch Suffigierung abgeleitete Wortformen betrachtet, nicht jedoch durch Präfigierung abgeleitete Wortformen.

Werden verschiedene Wurzeln zusammengesetzt, spricht man von „Komposition“. Bei den erhaltenen Worten handelt es sich demgemäß um „Komposita“. Komposita sind in der deutschen Sprache sehr häufig [MÖBIUS 2001]. Leider wird dadurch aber die Zerlegung von Worten in ihre Grundformen erschwert¹², für die so genannte „Kompositazerlegung“ sind spezielle Algorithmen erforderlich. Zu beachten sind dabei auch zwischen den einzelnen Bestandteilen eingefügte Buchstaben, die in der Regel die Aussprache erleichtern. Man spricht dabei von „Fugenelementen“ [MEIBAUER ET AL. 2002]. Führt man eine Kompositazerlegung durch, so kann man die Bestandteile der Komposita, die „Konstituenten“, bestimmen. Ferner können Beziehungen zwischen Komposita und den zu ihren Konstituenten verwandten Begriffen erschlossen werden¹³. Den rechten Konstituenten eines Kompositums bezeichnet man auch als dessen Kopf. Dieser bestimmt u.a. die Wortart. Bei den Komposita unterscheidet man im Wesentlichen Determinativkomposita und Kopulativkomposita. Bei Determinativkomposita wird die Bedeutung des rechten Elements durch das linke modifiziert [MEIBAUER ET AL. 2002], eine Eigenschaft die sich in der automatischen Sprachverarbeitung nutzen lässt. Ein Beispiel für ein Determinativkompositum im Bauwesen wäre „Betonstahl“, also ein Stahl, der durch den Zusatz Beton genauer bestimmt wird. In einem Kopulativkom-

12 Dies erkannte bereits Mark Twain 1880 in seinem Aufsatz „Die schreckliche deutsche Sprache“: „Das Wörterbuch muss irgendwo eine Grenze ziehen – also wird diese Art Worte ausgelassen. [...] Es handelt sich um zusammengesetzte Worte ohne Bindestriche. Die einzelnen Worte, die für ihren Bau benutzt wurden, stehen im Wörterbuch, aber sehr sehr verstreut; man kann einem Wort nach dem anderen hinterher jagen, und zu guter Letzt die Bedeutung verstehen, aber es ist ein ermüdendes und nervtötendes Geschäft.“

13 durch die Signifikanz des gemeinsamen Auftretens des Kompositums und der zu dessen Konstituenten verwandten Begriffen, vgl. Abschnitt 2.3.2

positum stehen beide Konstituenten gleichberechtigt auf derselben semantischen Ebene¹⁴ [KÖNIG 2001].

Im Rahmen der vorliegenden Arbeit sollen die in der Inhaltsbasis vorhandenen Terme soweit möglich auf ihre Stammformen zurückgeführt werden. Es liegen jedoch einschränkende Faktoren vor, die dieser Rückführung Grenzen setzen:

- Die die Stammformen enthaltende Datenbank soll zur Aufnahme neuer Wissenseinheiten nicht manuell ergänzt werden müssen, d.h., es kann für jedes neue Wort nur nachgesehen werden, ob eine – durch bestimmte Regeln zu bestimmende – Stammform vorhanden ist.
- Der als Stammform gespeicherte Term muss nicht der tatsächliche Wortstamm im sprachwissenschaftlichen Sinne sein, die oben beschriebene Vorgehensweise wird den – unter Anwendung der erwähnten Regeln – kürzesten im Textkorpus enthaltenen Wortbestandteil als Stammform zurückliefern.

Wird keine Syntax- bzw. Semantikanalyse der analysierten Texte vorgenommen, so ist es für die Funktionalität eines Systems unerheblich, ob es sich um den tatsächlichen Stamm oder eine als solcher erkannte kürzeste Wortform handelt.

Der Prozess der Rückführung der Terme auf Ihren Wortstamm wird in der Fachsprache des Information-Retrieval als „Stemming“ bezeichnet. Stemming ist in der deutschen Sprache schwieriger als z.B. im Englischen. Dies liegt einerseits daran, dass die gebeugten Formen der Worte eine größere Vielfalt aufweisen und andererseits an den komplexeren Regeln zur Bildung der gebeugten Formen.

Die Strukturierung von Sätzen nach grammatischen Regeln wird als Syntax bezeichnet (vgl. hier und im Folgenden [MEIBAUER ET AL. 2002]). Durch eine Analyse der Syntax lassen sich prinzipiell Aussagen über den Inhalt eines Satzes gewinnen. Dies ist eine komplexe Problematik, die außerdem den Einsatz eines Lexikons aller vorkommenden Wortformen und ihrer Bedeutung erfordert. Die Anpassung eines solchen Lexikons bei der Einbindung neuer Wissensdomänen – dem zu einem bestimmten Fachgebiet vorhandenen Wissen – würde immer einen nicht unerheblichen Anteil menschlicher Interaktion erfordern. Der Schwerpunkt der vorliegenden Arbeit liegt aber gerade in der weitestgehend automatischen Verarbeitung von Wissenseinheiten. Deshalb wird hier eine Syntaxanalyse nicht weiter verfolgt.

Ein weiterer wichtiger Aspekt in der automatischen Wissensverarbeitung ist die Semantik. Diese beschäftigt sich mit der Bedeutung sprachlicher Ausdrücke. Man unterscheidet hauptsächlich die

¹⁴ Kopulativkomposita in der Fachsprache des Bauingenieurwesens sind z.B. das Substantiv „Plattenbalken“ oder das Adjektiv bzw. Adverb „linearelastisch-idealplastisch“, das aber wegen des Bindestrichs eigentlich gar kein Kompositum ist. Die Konstituenten dieses Kopulativkompositums („linearelastisch“ und „idealplastisch“) sind ihrerseits wiederum Determinativkomposita.

Wort- und die Satzsemantik. Die Bestimmung semantischer Verhältnisse zwischen Worten ist Gegenstand der vorliegenden Arbeit. Semantische Verhältnisse („Relationen“) lassen sich in paradigmatische und syntagmatische unterteilen, wobei die paradigmatischen Relationen die Beziehungen gegeneinander austauschbarer Worte beschreiben und die syntagmatischen die Satzsemantik charakterisieren [KÖNIG 2001]. Die wichtigsten Beziehungen zwischen Worten einer Wortart sind nach MEIBAUER ET AL. 2002:

- *Synonymie*: Worte, die dieselbe Bedeutung haben. Man spricht von echten Synonymen, wenn die Worte in jedem Kontext austauschbar sind. Echte Synonyme sind in allen Sprachen selten. Beispiele für Synonyme im Bereich der Werkstoffe im Bauwesen sind z.B. „Anmachwasser“ und „Zugabewasser“.
- *Antonymie*: Antonyme Ausdrücke bezeichnen ein Paar von Gegensätzen. Zwischen diesen Gegensätzen kann es aber durchaus noch Abstufungen geben. Beispiele für Antonyme sind „kalt“ und „heiß“, eine Zwischenstufe wäre in diesem Fall „warm“.
- *Heteronymie*: Heteronyme Worte weisen von der Bedeutung her Gemeinsamkeiten auf, stehen aber für unterschiedliche Aussagen, z.B. „blau“, „rot“ und „grün“.
- *Hyponymie*: Hyponymie bezeichnet eine hierarchische Ordnung zwischen Worten. Die übergeordneten Begriffe werden als Hyperonyme, die untergeordneten als Hyponyme bezeichnet. Ein Beispiel ist das Hyponym „Spannstahl“ zum Hyperonym „Stahl“. Hyponyme zum selben Oberbegriff werden als Kohyponyme bezeichnet, z.B. „Spannstahl“ und „Betonstahl“.
- *Homonymie*: Die Homonymie bezeichnet eine Mehrdeutigkeit desselben Begriffs, z.B. „Bank“ für ein Sitzmöbel und ein Geldinstitut.
- *Polysemie*: Polysemie ist eine spezielle Form der Homonymie, hierbei kann aber der Begriff auf dieselbe Bedeutung zurückgeführt werden, z.B. „Oper“ für das Musikstück und das Gebäude.

2.3.2 Computerlinguistik

Die Computerlinguistik beschäftigt sich mit der automatischen Erfassung, Verarbeitung und Analyse von Texten. Insbesondere der Zweig des Text-Mining ist im vorliegenden Zusammenhang interessant. Dabei werden aus normalen, d.h. nicht speziell annotierten Texten der geschriebenen Sprache Informationen über die Inhalte der Texte gewonnen. Im Gegensatz zu einer Datenbank enthalten Texte Daten in unstrukturierter Form [HEYER ET AL. 2006]. Im betriebswirtschaftlichen

Zweig des Wissensmanagements werden Methoden der Computerlinguistik und des Text-Mining z.B. angewendet, um Fakten, Namen und Daten aus Texten extrahieren zu können.

Man unterscheidet in der Computerlinguistik lexikonbasierte und textstatistische Verfahren. Lexikonbasierte Verfahren benötigen Wörterbücher bzw. Thesauren, die i.d.R. manuell erstellt werden, um Aussagen zu Beziehungen zwischen Worten oder Texten zu treffen. Viele Informationen aus Texten lassen sich jedoch auch mittels textstatistischer Berechnungen gewinnen. Dabei werden vor allem die Häufigkeiten bestimmter Worte bestimmt. Diese lassen sich mit anderen Häufigkeiten, z.B. mit den Häufigkeiten derselben Worte in der Standardsprache oder mit denjenigen anderer Worte, vergleichen. Auch das gemeinsame Auftreten von Worten kann Rückschlüsse auf Beziehungen zwischen Worten liefern.

Texte lassen sich hinsichtlich ihrer Aufbereitung für die Nutzung mittels Text-Mining nach HEYER ET AL. 2006 in drei Stufen charakterisieren:

1. Originaldokumente im ursprünglichen Speicherformat
2. Zerlegung in Sätze, nur noch reiner Text
3. Verknüpfung der Wörter mit Informationen zu den Wörtern

Um Beziehungen zwischen Worten untersuchen zu können, muss man zunächst zwischen „syntagmatischen“ und „paradigmatischen“ Relationen zwischen den Worten unterscheiden [HEYER ET AL. 2006] [KÖNIG 2001]. Eine syntagmatische Relation zwischen zwei Begriffen liegt vor, wenn sie im selben „lokalen Kontext“ vorkommen, also beispielsweise im selben Satz stehen. Der globale Kontext eines Wortes ist die Menge aller Worte, die in einer statistisch signifikanten Anzahl gemeinsam mit dem Wort auftreten. Sind sich die globalen Kontexte zweier Worte hinreichend ähnlich, dann stehen diese Worte in paradigmatischer Relation.

Worte, die entweder in einer syntagmatischen oder in einer paradigmatischen Relation zueinander stehen, weisen eine semantische Relation auf (vgl. hier und im Folgenden [HEYER ET AL. 2006]). Diese semantischen Relationen können bei benachbarten Wortformen z.B. Kategorien, Maßeinheiten, Modifizierungen oder Veränderungen sein. Betrachtet man den globalen Kontext, so lassen sich u.a. auch Ober- und Unterbegriffe oder Synonyme identifizieren.

Treten zwei Begriffe in einem lokalen Kontext gemeinsam auf, so spricht man von einer Kookkurenz¹⁵. Unterschieden wird dabei zwischen Kookkurenzen benachbarter Worte („Nachbarkookkurenzen“) und Kookkurenzen von Worten, die in demselben Satz stehen („Satzkookkurenzen“). Aus den Nachbarkookkurenzen können eher anti-symmetrische Relationen abgeleitet werden, aus

¹⁵ In der Literatur wird synonym auch der Begriff „Kollokation“ verwendet. Dieser ist in der Sprachwissenschaft jedoch auch für feststehende Phrasen gebräuchlich (vgl. [MEIBAUER ET AL. 2002]). Aus diesem Grund wird in der vorliegenden Arbeit im Kontext gemeinsam auftretender Begriffe ausschließlich der Begriff „Kookkurenz“ benutzt.

den Satzkookkurrenzen eher symmetrische Relationen [HEYER ET AL. 2001]. Mit symmetrischen Relationen sind solche gemeint, die in beiden Richtungen gleich sind, also z.B. zwischen Synonymen. Anti-symmetrische Relationen finden sich z.B. zwischen Hyperonymen und Hyponymen sowie zwischen Adjektiven und Substantiven bzw. Adverbien und Verben.

Prinzipiell bildet jedes Wortpaar, das irgendwann einmal in einem Satz auftritt, eine Kookkurrenz, aber nicht jede Kookkurrenz stellt auch eine sinnvolle semantische Relation dar. Beispielsweise besteht in dem Satz „*Stahlkorrosion tritt auf, wenn Sauerstoff und Feuchtigkeit vorhanden sind, und Kunststoff altert unter UV-Licht.*“ keine semantische Beziehung zwischen „Stahlkorrosion“ und „UV-Licht“. Also muss ein Maß für die Stärke der semantischen Verknüpfung bestimmt werden, das berücksichtigt, wie oft Paare von Worten gemeinsam auftreten. Dies ist das „Signifikanzmaß“ der Kookkurrenz zweier Terme A und B. Es wird bei HEYER ET AL. 2006 als $\text{sig}(A,B)$ angegeben:

$$\text{sig}(A,B) = \frac{-\log\left(1 - e^{-x} \cdot \sum_{i=0}^{k-1} \frac{1}{i!} \cdot x^i\right)}{\log(n)} \quad (2.6)$$

Dabei stellen a und b die Anzahl der Sätze bzw. Paarungen benachbarter Worte dar, die Term A bzw. B enthalten, k ist die Anzahl der Sätze bzw. Paarungen benachbarter Worte, die beide Terme enthalten, n die Anzahl aller Sätze bzw. Paarungen benachbarter Worte des betrachteten Textkorpus. x bezeichnet den Ausdruck ab/n . Das Signifikanzmaß soll im Folgenden in Anlehnung an HEYER ET AL. 2006 hergeleitet werden.

Betrachtet werden die Sätze im Textkorpus unter dem Aspekt, ob die Worte A und B darin gemeinsam auftauchen. Eintreten können die beiden sich gegenseitig ausschließenden Ereignisse „gemeinsames Auftreten der Worte“ mit der Wahrscheinlichkeit p und „kein gemeinsames Auftreten der Worte“ mit der Wahrscheinlichkeit $1 - p$. Die Zufallsvariable k als Anzahl des gemeinsamen Auftretens der beiden Worte folgt damit einer Binomialverteilung [PAPULA 1994]. Die Prüfung, ob beide Worte in einem Satz vorkommen, wird für alle n Sätze durchgeführt. Als Mittelwert der Verteilung ergibt sich damit

$$\mu = n \cdot p \quad (2.7)$$

Die Wahrscheinlichkeit p setzt sich aus den beiden voneinander unabhängigen Wahrscheinlichkeiten p_A und p_B für das Auftreten der einzelnen Worte A und B in einem Satz zusammen. Mit der Häufigkeit a des Wortes A und der Häufigkeit b des Wortes B im Textkorpus aus n Sätzen gilt:

$$p = p_A \cdot p_B = \frac{a}{n} \cdot \frac{b}{n} \quad (2.8)$$

Kookkurenzen von beliebigen Worten treten per Definition in jedem Satz auf, das gemeinsame Auftreten zweier bestimmter Worte in einem Satz ist jedoch selten. Damit bietet sich die Beschreibung in Form einer Poisson-Verteilung an [WEBER 1992]. Diese stellt hier den Sonderfall einer Binomialverteilung für $n \rightarrow \infty$ und $p \rightarrow 0$ dar¹⁶. Für die Poisson-Verteilung gilt:

$$f(x, \lambda) = \frac{\lambda^x}{x!} \cdot e^{-\lambda} \quad (2.9)$$

Der Parameter λ stellt gleichzeitig den Mittelwert μ der Poisson-Verteilung dar [WEBER 1992], folglich gilt:

$$\lambda = n \cdot p = n \cdot p_A \cdot p_B = \frac{a \cdot b}{n} \quad (2.10)$$

Berechnet werden soll mit dieser Poisson-Verteilung jetzt die Wahrscheinlichkeit $P(X \geq k)$, dass beide Wortformen mindestens k-mal gemeinsam in einem Satz vorkommen. Diese Wahrscheinlichkeit ist äquivalent mit $1 - P(X < k)$. Der Term $P(X < k)$ lässt sich schreiben als

$$P(X < k) = \frac{\lambda^0}{0!} \cdot e^{-\lambda} + \frac{\lambda^1}{1!} \cdot e^{-\lambda} + \frac{\lambda^2}{2!} \cdot e^{-\lambda} + \dots + \frac{\lambda^{k-2}}{(k-2)!} \cdot e^{-\lambda} + \frac{\lambda^{k-1}}{(k-1)!} \cdot e^{-\lambda} \quad (2.11)$$

bzw.

$$P(X < k) = e^{-\lambda} \cdot \sum_{i=0}^{k-1} \frac{\lambda^i}{i!} \quad (2.12)$$

Für $P(X \geq k)$ gilt damit

$$P(X \geq k) = 1 - e^{-\lambda} \cdot \sum_{i=0}^{k-1} \frac{\lambda^i}{i!} \quad (2.13)$$

Nach HEYER ET AL. 2006 wird der Logarithmus dieser Wahrscheinlichkeit verwendet, da diese i.d.R. sehr klein ist. Der Logarithmus ist wegen $P(X \geq k) < 1$ negativ und wird deshalb noch mit -1 multipliziert. Um die Wahrscheinlichkeit auf die Größe des Textkorpus zu beziehen, wird anschließend noch durch den Logarithmus der Anzahl der Sätze dividiert. Somit gilt

¹⁶ Dies ist nach PAPULA 1994 zulässig, wenn $n \cdot p < 10$ und $n > 1500 \cdot p$ gilt. Für den WiBA-Net-Korpus ($n = 14.682$) ist die zweite Forderung in jedem Fall erfüllt, denn selbst für die häufigsten sinntragenden Begriffe „Stahl“ (693 Vorkommen) und „Beton“ (654 Vorkommen) gilt $14.682^3 > 1500 \cdot 693 \cdot 654$. Die erste Forderung ist für die erwähnten Begriffe nicht erfüllt, jedoch spätestens dann, wenn man nur die nicht ausgeschlossenen Worte ab dem hundertsten Rangplatz betrachtet. Diese haben die Häufigkeiten 237 (100. Platz) bzw. 235 (101. Platz). Für diese Begriffe gilt also $n \cdot p = 237 \cdot 235 / 14682 = 3,79 < 10$. Die wenigen sinntragenden Terme auf den Rangplätzen 1 bis 100 fallen nicht ins Gewicht und werden im Folgenden mit derselben Formel berechnet.

$$sig(A, B) = \frac{-\log\left(1 - e^{-\lambda} \cdot \sum_{i=0}^{k-1} \frac{1}{i!} \cdot \lambda^i\right)}{\log(n)} \quad (2.14)$$

was mit Formel (2.6) äquivalent ist. Nach HEYER ET AL. 2001 vereinfacht sich diese Gleichung für $2x < k$ zu:

$$sig(A, B) = \frac{x - k \cdot \log(x) + \log(k!)}{\log(n)} \quad (2.15)$$

Im Wesentlichen ist das Signifikanzmaß der Logarithmus der Wahrscheinlichkeit, dass die Begriffe A und B tatsächlich zufällig k-mal im selben Satz auftreten. Je geringer diese Wahrscheinlichkeit ist und je höher damit die Wahrscheinlichkeit eines semantischen Zusammenhangs zwischen A und B ist, desto größer ist das Signifikanzmaß.

Das Signifikanzmaß stellt keinen Absolutwert dar. Es lässt sich also keine für alle Paarungen gültige Schranke angeben, ab der die jeweiligen Paarungen eine semantische Relation darstellen. Stattdessen wird der Einzelfall betrachtet oder mit relativen Größen gerechnet.

Um die Aussagekraft hinsichtlich der Relationen zwischen Worten zu verbessern, können Kookkurenzen auch iterativ bestimmt werden. Dabei werden die in einem Textkorpus gefundenen Kookkurenzen selbst wieder als „Sätze“ betrachtet und nochmals analysiert, wobei sich häufig dieselben Relationen (z.B. Kohyponymie) zwischen allen so bestimmten Kookkurenzen feststellen lassen [HEYER ET AL. 2001].

Die in Abschnitt 2.2.3.1 vorgestellten Ähnlichkeitsmaße lassen sich durch textstatistische Methoden noch weiter verbessern. Auf diese Weise kann der Einfluss der Häufigkeit einzelner Worte auf die Ähnlichkeit von Dokumenten berücksichtigt werden. Grundannahme ist die Überlegung, dass sehr häufig vorkommende Begriffe nur eine geringe Diskriminierungswirkung haben und auch die Ähnlichkeit zweier Dokumente nicht hauptsächlich auf sehr häufigen Worten beruhen sollte¹⁷.

Zunächst wird der Anteil des untersuchten Begriffs an den Worten im Text bestimmt. Nach HEYER ET AL. 2006 lässt sich die normalisierte Termfrequenz $nf_{i,m}$ als relative Häufigkeit wie folgt schreiben:

$$nf_{i,m} = \frac{f_{i,m}}{\sum_{t_j \in d_m} f_{j,m}} \quad (2.16)$$

¹⁷ Bei zwei Dokumenten, in denen jeweils zehn mal der Begriff „die“ (9315 mal im WiBA-Net-Korpus enthalten) vorkommt, kann nur eine viel unpräzisere Vermutung zu ihrer Ähnlichkeit vorgenommen werden als bei zwei Dokumenten, die z.B. je einmal den Begriff „Polykondensation“ (22 mal enthalten) enthalten.

Der Index m bezeichnet hier und im Folgenden das Dokument, der Index i den betrachteten Term. $f_{i,m}$ steht für die Frequenz des Terms t_i im Dokument d_m . Der Nenner steht für die Summe aller Terme im Dokument d_m .

Durch die Verwendung der Inversen Dokumentfrequenz (IDF) können selten vorkommende Terme in der Gewichtung bevorzugt werden [LEWANDOWSKI 2005]. Diese Bevorzugung lässt sich durch die Verwendung eines logarithmierten Wertes ein wenig abschwächen [FERBER 2003], sodass sich nach HEYER ET AL. 2006 schreiben lässt:

$$IDF_i = \log\left(\frac{m}{d:t_i \in d}\right) \quad (2.17)$$

m steht hier für die Menge der Dokumente, der Nenner bezeichnet die Menge aller Dokumente, in denen der Term t_i vorkommt.

Die Gewichtung eines Terms t_i in einem Dokument d_m lässt sich nach HEYER ET AL. 2006 als Produkt der normalisierten Termfrequenz und der Inversen Dokumentfrequenz darstellen. Es gilt für den Term t_i im Dokument d_m :

$$w_{i,m} = n f_{i,m} \cdot IDF_i \quad (2.18)$$

Das Kosinusmaß nach Formel (2.1) lässt sich unter Einbeziehung dieser Gewichtung nach HEYER ET AL. 2006 (vgl. auch [FERBER 2003]) wie folgt schreiben:

$$\text{sim}_{\cos}(\vec{d}_i, \vec{d}_j) = \frac{\sum_{k=1}^n (w_{k,i} \cdot w_{k,j})}{\sqrt{\sum_{k=1}^n (w_{k,i})^2} \cdot \sqrt{\sum_{k=1}^n (w_{k,j})^2}} \quad (2.19)$$

Will man nicht die Ähnlichkeit von Dokumenten untereinander bestimmen, sondern die zweier Begriffe quantifizieren, so kann man die Ähnlichkeit zwischen den Kookkurentenvektoren der beiden Begriffe bestimmen. Einen Begriff, der sowohl zu einem Wort A als auch zu einem Wort B eine Kookkurenz darstellt, bezeichnet man als „Link“ zwischen diesen Begriffen, die sich ergebende Ähnlichkeit zwischen den Begriffen A und B als „indirekte Assoziation“, die als Verknüpfungsstärke zwischen den Begriffen interpretiert werden kann [ACKERMANN 2000].

Das Verfahren lässt sich gut zur Suchterweiterung nutzen. Dabei muss beachtet werden, dass die berechneten Signifikanzmaße und Ähnlichkeiten nicht denselben Maßstab haben. So kann die Signifikanz der Kookkurenz zweier Begriffe A und B beispielsweise 1,523 betragen. Das auf die beschriebene Weise bestimmte Kosinusmaß zwischen dem Begriff A und einem weiteren Begriff C

betrage z.B. 0,847. Eine „Rangordnung“ der Ähnlichkeit von B oder C zu A lässt sich aus diesen Zahlen jedoch nicht entnehmen.

3 Lernnetzwerk „Werkstoffe im Bauwesen“ (WiBA-Net)

3.1 Vorbemerkung

Das Lernnetzwerk „Werkstoffe im Bauwesen und in der Architektur“ (WiBA-Net) ist das Ergebnis eines umfangreichen Forschungsprojektes. Es wurde im Rahmen der vom BMBF geförderten Forschungsinitiative „Neue Medien in der Bildung“ ab 2001 als hochschul- und fachübergreifendes Projekt entwickelt. Die wissenschaftliche Leitung lag bei Herrn Prof. Grübl von der TU Darmstadt. An dem Verbundprojekt waren als inhaltlich Verantwortliche Professoren und wissenschaftliche Mitarbeiter aus dem Bauingenieurwesen und der Architektur an verschiedenen Hochschulen beteiligt. Die Verantwortung für die Technik und die Pädagogik lag bei der TU Darmstadt¹⁸.

In diesem Kapitel wird das Gesamtprojekt „WiBA-Net“ beschrieben. Seine Kenntnis ist wichtig für die Einordnung des in den folgenden Kapiteln vorgestellten Vorschlags zur automatischen Erstellung von Lerneinheiten. Dieser Vorschlag stellt zum einen eine Erweiterung der Funktionalitäten des WiBA-Net dar. Zum anderen kann er aber auch auf die Inhalte anderer Wissensgebiete angewendet werden, wenn diese textbasiert sind und sich entsprechend aufbereiten lassen. In Abschnitt 3.5 wird außerdem explizit mein eigenständiger Beitrag zum WiBA-Net vorgestellt.

3.2 Konzept des WiBA-Net

Die Inhalte des WiBA-Net entsprechen zum einen, wie bereits erwähnt, den im Hochschullehrer-memorandum [REINHARDT 2000] enthaltenen inhaltlichen Anforderungen an das Fach „Werkstoffe im Bauwesen“. Abgedeckt werden danach insbesondere die Themenschwerpunkte „Allgemeine Grundlagen“, „Metallische Werkstoffe“, „Organische Werkstoffe“ und „Mineralische Werkstoffe“.

Die Inhalte sind von den Hochschullehrern festgelegt worden, die an deutschsprachigen Hochschulen das Fach „Werkstoffe im Bauwesen“ vertreten. Sie werden von ihnen allgemein als Mindestlehrinhalte angesehen, welche im Grundstudium vermittelt werden sollten. Bei der Umstellung des Studiums im Zuge des Bologna-Prozesses sind diese Inhalte Gegenstand der Ausbildung zum Bachelor geworden. Jedem Hochschullehrer ist es aber freigestellt, eigene Schwerpunkte zu setzen. Zum anderen werden die oben genannten Themenschwerpunkte ergänzt um das weitere Thema „Bezug zur Konstruktion“ sowie um Gesichtspunkte der Architektur und der Werkstoffmechanik. Jedes Einzelthema wurde von jeweils einem der beteiligten Institute bearbeitet.

Zielgruppe des Lernnetzwerks sind alle Studierenden des Bauingenieurwesens und der Architektur. Das WiBA-Net wurde dabei von Anfang an als eine Ergänzung zur Vorlesung und nicht als ein

¹⁸ Eine Übersicht der Projektbeteiligten findet sich im Anhang E.

Ersatz für diese konzipiert. Der Einsatz der Inhalte und Funktionen des WiBA-Net kann entsprechend der Anforderungen des jeweiligen Dozenten variabel gestaltet werden.

WiBA-Net kann sowohl in der Vorlesung („Präsenzveranstaltung“) selbst als auch online nach dem Konzept des „Blended Learning“ zur Vor- und Nachbereitung der Themen einzelner Vorlesungen genutzt werden. Durch die vollständige inhaltliche Abdeckung des Hochschullehrermemorandums ist den Studierenden auch das komplette Erarbeiten versäumter oder ausgefallener Vorlesungen möglich. Für ein reines Online-Studium der gesamten Lehrveranstaltung ist das Lernnetzwerk primär nicht gedacht. Generell kann das WiBA-Net durch seinen flexiblen Aufbau aber von jedem Hochschullehrer nach dessen individuellem Ermessen genutzt werden.

Heute steht der Einsatz von WiBA-Net allen Hochschullehrern des Fachs „Werkstoffe im Bauwesen“ frei, auch den nicht an der Entwicklung beteiligten. Nach dem Auslaufen der Projektförderung wurde ein Förderverein zur Betreuung und Pflege des WiBA-Net gegründet.

3.3 Funktionalitäten des WiBA-Net

Alle Funktionen des WiBA-Net stehen über das Internet zur Verfügung [GRÜBL ET AL. 2004]. Sie lassen sich zu den Gruppen der kompetenzvermittelnden, kompetenzvertiefenden, kommunikativen und additiven Komponenten zusammenfassen¹⁹. Einen Überblick über diese Komponenten bietet Tabelle 3.1. Den inhaltlichen Schwerpunkt des WiBA-Net bilden die kompetenzvermittelnden Komponenten. Diese werden im Folgenden ausführlicher als die übrigen Komponenten beschrieben, da auf ihnen der im Rahmen der vorliegenden Arbeit entwickelte Vorschlag zur automatischen Generierung von Lerneinheiten beruht.

Die kompetenzvermittelnden Komponenten lassen sich grundsätzlich in Funktionen für das Online-Studium und solche für die Präsenzlehre unterteilen. Außerdem werden ein Bereich für Hochschullehrer und einer für Studierende unterschieden. Als verantwortliche Autoren bzw. Produzenten von Inhalten sind nur die Hochschullehrer zugelassen. Sie haben die Möglichkeit, multimediale Elemente („Assets“), Texte in Form von HTML-Seiten, („Wissenseinheiten“), sowie daraus zusammengestellte Lerneinheiten („Lehrpfade“) auf den Server hoch zu laden und in einem hierarchisch strukturierten Bereich, dem „InfoPool“, abzulegen.

¹⁹ Der Begriff „Kompetenz“ bezeichnet in diesem Zusammenhang in erster Linie die Vermittlung und Vertiefung von Fachkompetenz.

Tabelle 3.1: Übersicht über die Funktionalitäten des WiBA-Net

Kompetenzvermittelnde Komponenten		Kompetenzvertiefende Komponenten	Kommunikative Komponenten	Additive Komponenten
Grundlage: WiBA-InfoPool	E-Learning - Seiten / Wissens-einheiten - Lehrpfade - Lehrgänge - Kurse	Übungsmodul / Lernkontrollen	Online-Sprechstunde	Bezug zu Regelwerken
	Hochschullehrerfunktionen - Assets	Aufgabenpool	Chatroom	externe elektronische Information
		Virtuelles Praktikum	Forum	Literatur

Alle textlichen Inhalte bzw. Wissensseinheiten werden in Form von HTML-Seiten vorgehalten²⁰. In der Regel entspricht eine Seite dem Inhalt, welcher insgesamt auf einer Bildschirmseite dargestellt werden kann, ohne dass „gescrollt“ werden muss. In die Seiten sind die jeweils benötigten Multimedia-Elemente eingebunden. Das wichtigste Prinzip bei der Erstellung von Seiten ist, auf jeder Seite nur einen Inhalt zu behandeln. Diese Inhalte sind kontextfrei, also ohne direkten textlichen Bezug auf andere Seiten bzw. Wissensseinheiten. Links auf andere Seiten sind in den Wissensseinheiten ebenfalls nicht enthalten. Die Seiten werden erst bei der Erzeugung von Lehrpfaden in einen Kontext zueinander gestellt. Diese Vorgehensweise gewährleistet eine hohe Wiederverwendbarkeit der erstellten Inhalte.

Zum Wiederauffinden der Inhalte und um ihre Wiederverwendbarkeit sicherzustellen, verfügen die Inhalte des WiBA-Net über Metainformationen nach dem SCORM-Standard [ADL 2008]. Diese umfassen alle vom SCORM-Standard festgelegten Elemente. Sie können beim Hochladen der Dateien teilweise automatisch ausgefüllt werden (z.B. Größe oder Typ der Datei). Die zur Identifizierung bei einer Suche dienenden Elemente werden von den Autoren manuell eingetragen. Dies sind vor allem der Titel des Elements, eine Beschreibung und Stichworte zu seinem Inhalt. Die im WiBA-Net implementierte Suchfunktion arbeitet auf der Basis dieser Metadaten.

Den Studierenden werden die Wissensseinheiten im Regelfall im Rahmen der Lehrpfade (vgl. Abschnitt 3.4.1) angezeigt. Ein Zugang über den InfoPool oder die Suchfunktion ist nur im Ausnahmefall vorgesehen. Es wird davon ausgegangen, dass die Studierenden nur geringe Vorkenntnisse haben und ihnen deswegen zusammengehörige Informationen am besten in ihrem jeweiligen Kontext angeboten werden. Multimedia-Elemente sind für Lerner standardmäßig nur zugänglich, wenn sie in HTML-Seiten mit erläuterndem Text eingebunden sind. Als eigenständige Dateien stehen sie nur für Hochschullehrer zum Einsatz in Präsenzveranstaltungen zur Verfügung. Lehrpfade werden nach übergeordneten Themen zu „Lehrgängen“ zusammengefasst. Mehrere Lehrgänge bilden zu-

²⁰ Eine Liste aller im WiBA-Net enthaltenen Wissensseinheiten findet sich im Anhang F.

sammen wiederum einen „Kurs“. Titel bzw. Themen der Kurse sind u.a. die einzelnen Baustoffe, aber auch übergreifende Themen. So gibt es z.B. Kurse mit den Themen „Holz“, „Beton“ oder „Dauerhaftigkeit“.

Die „Hochschullehrerfunktionen“ (vgl. Abschnitt 3.5) ermöglichen es den Lehrenden, Materialien (in der Regel Bildmaterial) für eine bevorstehende Vorlesung auszuwählen, zusammenzustellen und in der Präsenzveranstaltung anzuzeigen. Die Anzeige erfolgt dabei in Form einer Präsentation analog zu Microsoft Powerpoint.

Die kompetenzvertiefenden Komponenten beinhalten Elemente, die die Anwendung des erlernten Wissens fördern und das Behalten des Gelernten erleichtern sollen. Im Aufgabenpool können die Studierenden Skripte zum Selbststudium, Materialien aus der Vorlesung, Aufgabenstellungen für anstehende Präsenzübungen und Praktika sowie alte Klausuraufgaben zur Prüfungsvorbereitung herunterladen. Die Studierenden können ihren Lernerfolg mit Hilfe von Lernkontrollen, so genannten „Tests“, überprüfen. Ihnen stehen am Ende jeder Lerneinheit und zusätzlich lerneinheitsübergreifend Testaufgaben zur Verfügung. Diese wurden mit der Software *Perception* der Firma *Questionmark* realisiert. Es handelt sich um Fragen, die den Inhalt der Lerneinheiten umfassen. Zur Verfügung stehen Single- und Multiple-Choice-Fragen, Rechenaufgaben, Zuordnungen, Lückentexte und Freitexte. Die Auswertung erfolgt automatisch. Das Ergebnis der Online-Tests dient den Lernern zur Selbstkontrolle und hat keinen Einfluss auf die zur Lehrveranstaltung gehörende Prüfung.

Der wesentliche Bestandteil der kommunikativen Komponenten ist ein Chatroom mit integriertem Whiteboard. Er ermöglicht die Kommunikation der Lerner untereinander und mit den Dozenten bzw. Tutoren. Das Whiteboard erlaubt dabei das Hochladen von Grafikdateien in den Hintergrund einer für alle Nutzer sichtbaren Zeichenfläche. Diese kann auch ohne Hintergrundbild genutzt werden. Im Rahmen regelmäßiger Online-Sprechstunden können Fragen und Verständnisprobleme geklärt werden. Dabei kann die Zeichenfläche, mit oder ohne Hintergrundbild, von den Tutoren zur grafischen Erläuterung eingesetzt werden. Die asynchrone Kommunikation zwischen den Nutzern findet über ein Forum statt.

Zu den additiven Komponenten gehören eine Liste der verwendeten Normen und Regelwerke, eine Linksammlung und Literaturangaben. Den Hochschullehrern steht ferner eine anonyme Statistik aller Testergebnisse zur Verfügung. So können beispielsweise nach einer Vorlesung Testfragen zu dieser speziellen Veranstaltung angeboten werden. Der Dozent kann so einen Eindruck vom Lernerfolg seiner Studierenden bzw. seines eigenen Lehrerfolgs gewinnen. Für die Dozenten und Betreuer wird der Organisationsaufwand für die Durchführung der Lehrveranstaltung gesenkt, da viele Publikationsaufgaben über das WiBA-Net abgewickelt werden können. Termine und Ergebnisse von Tests und Klausuren werden ebenfalls über das WiBA-Net veröffentlicht.

3.4 Selbststudium mit Lehrpfaden

3.4.1 Konzept der Lehrpfade

Die im Regelfall²¹ von den Studierenden zu nutzende Lerneinheit des WiBA-Net ist der so genannte „Lehrpfad“. Das Konzept der Lehrpfade wurde im Rahmen des Forschungsprojekts „WiBA-Net“ entwickelt. Es handelt sich dabei um eine thematisch verwandte Menge von Wissenseinheiten, die in einer didaktisch sinnvollen Struktur zusammengefügt sind und die in der Regel sequentiell bearbeitet werden. Jeder Lehrpfad befasst sich nur mit einem abgegrenzten Thema, das nicht mehr sinnvoll in Unterthemen aufgegliedert werden kann. Ein Lehrpfad ist so konzipiert, dass seine Bearbeitungsdauer für einen Studierenden im Durchschnitt 25 Minuten beträgt. Die Bearbeitungsdauer einzelner Lehrpfade kann je nach Thematik aber zwischen 15 und 45 Minuten variieren. Auf diese Weise soll den Studierenden das Bearbeiten jeweils eines Lehrpfades auf einmal, d.h. ohne Pausen, ermöglicht werden.

Die Wissenseinheiten, die einen Lehrpfad bilden, können entweder linear hintereinander oder verzweigt angeordnet sein. Die verzweigte bzw. hierarchische Anordnung stellt dabei den Regelfall dar. Jede Wissenseinheit enthält genau einen Inhalt und hat keine direkten textlichen Bezüge zu anderen Wissenseinheiten. Der Kontext des Lehrpfads wird erst durch die Zusammenstellung der Wissenseinheiten zu Lehrpfadseiten erreicht. Dabei enthält der Lehrpfad nur Wissenseinheiten, also HTML-Seiten, niemals einzelne multimediale Elemente. Alle solchen Elemente werden zuvor in HTML-Seiten eingebettet, dabei mit einer Überschrift versehen und im Allgemeinen auch erläutert, sodass eine Wissenseinheit entsteht.

Wesentliches Prinzip ist die Untergliederung der Seiten eines Lehrpfads in „Hauptseiten“ und „Nebenseiten“ (vgl. Abbildung 3.1). Hauptseiten sind auf der obersten Ebene des Lehrpfads angesiedelt und enthalten grundlegende Informationen zu einem bestimmten Thema. Sie werden durch untergeordnete Nebenseiten zu vertiefenden Themen ergänzt. Die Nebenseiten sind normalerweise ein oder zwei Ebenen tief geschachtelt. Für die jeweils tiefere Ebene gilt analog, dass sie zur höheren Ebene vertiefende Informationen enthält. Der Nutzer kann den Lehrpfad entweder komplett durchlaufen oder sich durch Bearbeiten nur der Hauptseiten²² einen Überblick zu einem Thema verschaffen. Ein Lerner, der sich im Grundstudium befindet und keine Vorkenntnisse hat, sollte immer den gesamten Lehrpfad bearbeiten, da dieser vollständige und detaillierte Informationen zu dem vorliegenden Thema bietet.

21 Im Fall einer Suche werden dem Nutzer auch Wissenseinheiten, also einzelne Seiten, angezeigt. Diese können einzeln betrachtet werden.

22 Wahlweise können auch einzelne Ebenen der Nebenseiten mit einbezogen werden.

Ergänzend kommen zu den Haupt- und Nebenseiten noch „weiterführende Seiten“ hinzu. Diese gehören inhaltlich nicht zum eigentlichen Thema des Lehrpfads, stehen den Inhalten der Haupt- und Nebenseiten jedoch nahe und zeigen Bezüge zu anderen Themen bzw. Lehrpfaden auf. Da sie nicht zum Inhalt des Lehrpfads gehören, sind die auf ihnen behandelten Themen nicht Bestandteile des den Lehrpfad abschließenden Tests.

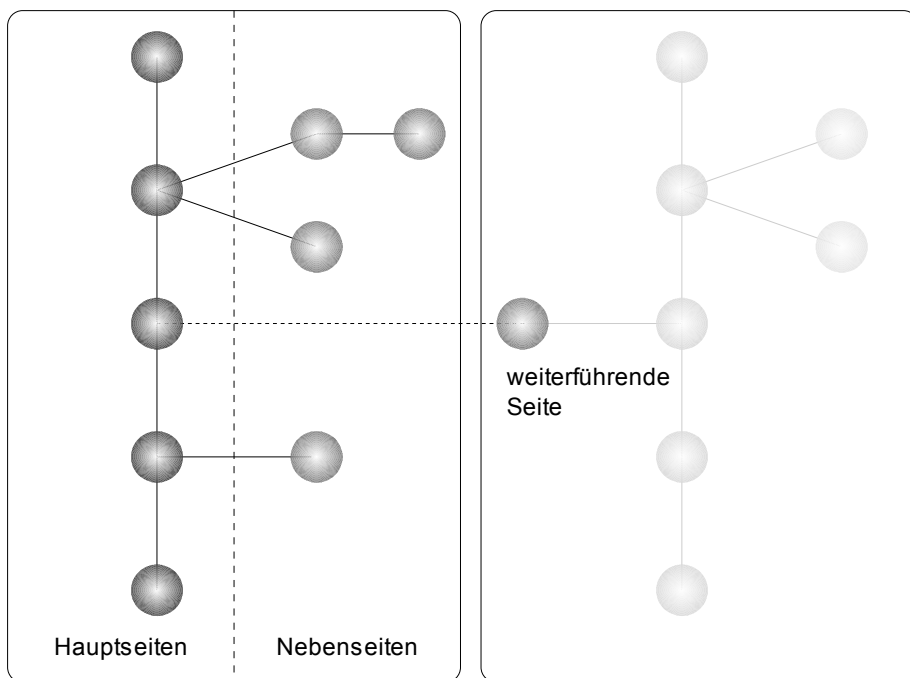


Abbildung 3.1: Schematische Darstellung eines Lehrpfads als Menge von Haupt-, Neben- und weiterführenden Seiten

Die Struktur des gesamten Lehrpfads ist für den Lerner zu jedem Zeitpunkt sichtbar. Auf diese Weise kann schnell auf die Informationen der anderen Seiten zugegriffen werden, eine lineare Abarbeitung der Lerneinheiten durch alleiniges Vor- und Zurückblättern wird nicht erzwungen. Außerdem ist der Lerner darüber informiert, welche Inhalte er schon bearbeitet hat und welche noch kommen. Auf diese Weise kann das auf der aktuellen Seite Gelernte jederzeit in den Kontext des Gesamtlehrpfads eingeordnet werden. Die Darstellung einer Seite aus einem Lehrpfad inkl. der Navigationselemente findet sich in Abbildung 3.2. Aufbau und Inhalte eines vollständigen WiBA-Net-Lehrpfads sind im nächsten Abschnitt dargestellt.

Der Planung von Lerninhalten sollten, wie in Abschnitt 2.1.2 beschrieben, die „Instruktionsereignisse“ nach Gagné zugrunde gelegt werden. In Anlehnung daran werden Lehrpfade immer durch die Nennung des Lehrziels eingeleitet und schließen mit einer Lernkontrolle ab. Das Lehrziel fasst den Inhalt des Lehrpfads zusammen, nennt die erreichbare Kompetenz nach dem Durcharbeiten des Lehrpfads und gibt die durchschnittliche Bearbeitungsdauer an. Die Lernkontrolle besteht aus au-

tomatisch ausgewerteten Testfragen zu den Inhalten des Lehrpfads. Verschiedene Typen von Testfragen sind möglich. Die Ergebnisse des Tests dienen den Studierenden lediglich zur Selbsteinschätzung, alle Ergebnisse werden danach nur anonymisiert gespeichert.



Abbildung 3.2: Seite aus dem WiBA-Net Lehrpfad „Konsistenzbestimmung“

Der didaktische Aufbau der Lehrpfade folgt im Allgemeinen einem deduktiven Ansatz, indem von allgemeinen Inhalten zu spezielleren Themen voran geschritten wird. Die Hauptseiten werden also in der Regel vom Anfang zum Ende eines Lehrpfads immer spezifischer. Da die Autoren frei in der Wahl ihrer didaktischen Methoden waren, weisen nicht alle Lehrpfade genau diesen Aufbau auf.

Im Projekt WiBA-Net wurde zur Erstellung und Anzeige der Lehrpfade das Programm *Learning Solution* der Firma SAP eingesetzt. Dieses zeichnet sich durch einen hohen Funktionsumfang aus und ist von Studierenden einfach zu bedienen. Es erfordert aber von den Autoren einen hohen zeitlichen Aufwand für die Einarbeitung und die Erstellung von Lehrpfaden. Ferner ist es erforderlich, die in den Lehrpfaden enthaltenen HTML-Seiten sowie die in diese eingebundenen Assets vor der Erstellung eines Lehrpfads separat auf den Server zu laden und die Seiten im Lehrpfad dort zu

referenzieren. Die „spontane“ Zusammenstellung vorhandener Wissenseinheiten zu neuen Lehrpfaden, die aktuell in der Vorlesung behandelte oder zu behandelnde Themen umfassen, ist nicht praktikabel.

3.4.2 Beispiel eines WiBA-Net Lehrpfads

In Tabelle 3.2 ist exemplarisch die Struktur²³ des WiBA-Net-Lehrpfads „Konsistenzbestimmung“ dargestellt. Die Seiten des Clusters „Versuche zur Konsistenzbestimmung“ sind zur Veranschaulichung im Anhang G als Screenshots abgebildet.

Tabelle 3.2: Aufbau des Lehrpfads „Konsistenzbestimmung“

Nr.	Hauptseiten	Nebenseiten	
1	Lehrziel		
2	Einführung Konsistenz		
3		Zeitraum der Konsistenzbeibehaltung	
4			Dauer der Konsistenzbeibehaltung
5	Konsistenzbereiche		
6		Erläuterungen zu den verschiedenen Konsistenzbereichen	
7			Sehr steife Konsistenz
8			Steife Konsistenz
9			Plastische Konsistenz
10			Weiche Konsistenz
11			Sehr weiche Konsistenz
12			Fließfähige und sehr fließfähige Konsistenz
13			Fließbeton
14			Konsistenzklassen
15			Konsistenzmaße
16			Beispiele für Konsistenzklassen, -maße und -bereiche
17	Konsistenzen für verschiedene Einsatzzwecke		
18		Anforderungen an die Frischbetonkonsistenz in Abhängigkeit von Transport, Einbau und Verdichtung	
19		Anforderungen an die Frischbetonkonsistenz in Abhängigkeit vom Bauteil	
20	Versuche zur Konsistenzbestimmung		
21		Bestimmung des Ausbreitmaßes	
22			Ausbreitmaßklassen
23			Vergleich der Konsistenzklassen

²³ Ursprünglich sind bei diesem Lehrpfad einige Nebenseiten mehreren Hauptseiten zugeordnet. Da dies nicht dem WiBA-Net-Konzept entspricht, wird an dieser Stelle jede Seite nur einmal dargestellt.

Nr.	Hauptseiten	Nebenseiten	
24		Bestimmung des Verdichtungsmaßes	
25			Verdichtungsmaßklassen
26		Bestimmung des Slumpmaßes	
27			Slumpmaßklassen
28		Bestimmung des Vébémaßes	
29			Vébémaßklassen
30	Test		

3.5 Entwicklung einer Software zur Unterstützung von Präsenzveranstaltungen

Das Lernnetzwerk WiBA-Net wurde, wie oben beschrieben, so konzipiert, dass es sowohl von den Studierenden zum Selbststudium, z.B. für die Vor- und Nachbereitungsphasen einer Vorlesungsreihe, herangezogen als auch vom Dozenten zur Unterstützung der Vorlesung genutzt werden kann. Es waren somit zwei Ansprüche gleichzeitig zu erfüllen, die sich nicht ohne weiteres vereinbaren lassen: Die in den Lehrpfaden verwendeten Wissensseinheiten enthalten zu einem großen Teil Fließtext und eignen sich damit nur bedingt für den Einsatz in einer Präsentation, da der enthaltene Text Gegenstand des mündlichen Vortrags des Dozenten ist. Dieser Vortrag wird üblicherweise um schriftliche Komponenten und zusätzliches gezeigtes Material ergänzt. Die schriftlichen Passagen werden meist entweder an der Tafel oder auf dem Overhead-Projektor entwickelt. Zusätzliche Materialien werden heute in der Regel in Form von Powerpoint-Folien gezeigt, aber auch der Overhead-Projektor kommt noch häufig zum Einsatz. Die Visualisierung von Sachverhalten mittels Bildern oder Diagrammen erlaubt es Lernern, Informationen besser zu verarbeiten und Lösungswege aufgezeigter Probleme leichter nachzuvollziehen [PFÄFFLI 2005].

Das Interesse des Dozenten besteht überwiegend an der Verwendung nicht-textlicher Medien²⁴. Für die Zusammenstellung der multimedialen Inhalte des WiBA-Net für die Präsentation in Präsenzveranstaltungen wurde das Programmmodul „Elsbeth“²⁵ entwickelt. Alle verwendbaren Bild- und Bewegtmedien des WiBA-Net sind – wie oben beschrieben – mit Metadaten nach dem SCORM-Standard ausgezeichnet, es kann also auf Titel, Stichworte und in vielen Fällen auch eine Kurzbeschreibung des Inhalts des Mediums zurückgegriffen werden. Auf diese Weise können die zum Thema einer bestimmten Vorlesung passenden Medien (Bilder, Diagramme, Animationen etc.) schnell gefunden und zusammengestellt werden. Die Verwendung der multimedialen Inhalte

24 Bilder (Zeichnungen und Fotos), Diagramme, Animationen, Videos

25 Akronym für „Elektronische Lernelementsortierung und -beschriftung – Erstellung teilautomatischer Hörsaalpräsentationen“

des WiBA-Net in Präsenzveranstaltungen erhöht den Wiedererkennungswert zwischen dem in der Vorlesung eingesetzten Material und den Selbstlernkursen des WiBA-Net.

Der Nutzer kann in der Vorlesung auf diese Zusammenstellung – bei entsprechender Ausstattung des Hörsaals auch online – zurückgreifen und die einzelnen Inhalte im Rahmen der Vorlesung den Studierenden über einen Beamer anzeigen. Für eine hohe Lerneffizienz ist es nach PFÄFFLI 2005 wichtig, auf den Kenntnisstand der Lerner einzugehen und Fragen und Verständnisprobleme der Studierenden auch in der Vorlesung zu berücksichtigen. Aus diesem Grund kann bei einer bestehenden Internetverbindung während einer laufenden Vorlesung vom Dozenten eine für das Auditorium nicht sichtbare Suchmaske genutzt werden. Diese erlaubt das Einfügen von aktuellen Materialien, deren Einsatz bei der Vorbereitung der Vorlesung nicht geplant war.

Bei Einsatz eines Tablet-PCs – bei dem mit einem speziellen Stift direkt auf dem Bildschirm geschrieben werden kann – oder eines Grafiktablets kann der Dozent einzelne Annotationen oder auch die gesamte Tafelanschrift innerhalb des Moduls „Elsbeth“ vornehmen, da direkt in die Folien geschrieben werden kann. Für die Entwicklung längerer Tafelanschriften können dabei an jeder Stelle der Präsentation weiße Folien eingefügt werden. Durch diese Vorgehensweise wird versucht, mediale Brüche, die aus der gleichzeitigen Verwendung einer Laptop-Präsentation und einer Tafel bzw. eines Overhead-Projektors resultieren, zu vermeiden. Die eingegebenen Schriftelemente werden bei jedem Folienwechsel automatisch gespeichert.

Die ggf. mit Kommentaren versehenen oder um Tafelanschriften ergänzten Inhalte können nach der Vorlesung den Studierenden über das WiBA-Net zur Verfügung gestellt werden. So können diese die Vorlesung in ihren wesentlichen Teilen nochmals selbst rekapitulieren. Für die Nachbereitung der Vorlesung können auch Verbindungen mit themenverwandten Lehrpfaden hergestellt werden. Fehlt während der Vorlesung eine Internetanbindung, so können keine Materialien vom Server nachgeladen werden. Das Modul wird dann wie eine herkömmliche Präsentationssoftware genutzt. Eventuelle Annotationen werden lokal gespeichert und können nach der Vorlesung auf den Server übertragen werden.

Durch die Nutzung des Software-Moduls „Elsbeth“ können die Studierenden sich besser auf den Vortrag des Dozenten konzentrieren und sich dazu eigene Notizen machen. Dies fördert das Behalten des Lernstoffs, da durch den Transfer des Gehörten in eigene Worte vom Lerner bereits neues Wissen im Sinne des Konstruktivismus generiert wird. Es muss nicht notwendigerweise die gesamte Tafelanschrift abgeschrieben werden.

3.6 Einsatz des WiBA-Net zur Vorlesungsunterstützung

Bei der universitären Lehre lässt sich nach DIRECTORATE-GENERAL 2005 zwischen der Vorbereitungsphase, der Präsenzphase und der Nachbereitungsphase einer Vorlesung unterscheiden. Das WiBA-Net unterstützt in allen drei Phasen sowohl Hochschullehrer als auch Studierende. Die jeweiligen Möglichkeiten der Unterstützung werden im Folgenden beschrieben. Eine Zusammenstellung der Phasen und der jeweils genutzten Funktionalitäten des WiBA-Net enthält Tabelle 5.1 in Abschnitt 5.1.1. Diese Tabelle beinhaltet zusätzlich die Anwendungsmöglichkeiten der im Rahmen dieser Arbeit vorgeschlagenen automatisch generierten Lehrpfade. Diese werden in Kapitel 5 im Detail vorgestellt.

Lehrpfade für die Vorbereitungsphase

Im Nachgang zur vorhergehenden Vorlesung wird den Studierenden das Thema der nächsten Vorlesung bekannt gegeben, sodass sie sich inhaltlich darauf vorbereiten können. Die Vorbereitung wird ihnen durch die zentrale Bereitstellung von Lehrpfaden zum Thema der nächsten Vorlesung im WiBA-Net erleichtert.

Der Dozent bereitet in der Vorbereitungsphase seinen Vortrag vor. Dieser kann durch eine Präsentation multimedialer Elemente ergänzt werden. Die Präsentation wird aus den im InfoPool des WiBA-Net vorhandenen multimedialen Elementen zusammengestellt.

Webbasierte Hörsaalpräsentationen für die Präsenzveranstaltung

Die Vorlesung als eigentliche Präsenzveranstaltung besteht vor allem aus dem Vortrag des Hochschullehrers. Dieser kann auf den in der Vorbereitungsphase von den Studierenden erarbeiteten Inhalten aufbauen. Der Vortrag wird je nach Thema der Vorlesung durch Medien (Grafiken, Diagramme) und eine Tafelanschrift ergänzt. Die aus dem InfoPool zusammengestellten Medien werden mit Hilfe des Programmmoduls „Elsbeth“ mit einem Beamer präsentiert. Die Tafelanschrift kann bei Verwendung eines Tablet PCs oder eines Grafiktablets innerhalb derselben Präsentation erfolgen. Sie wird zusammen mit den verwendeten Medien gespeichert.

Haben sich die Studierenden auf die Veranstaltung vorbereitet, ist auch eine Erarbeitung des eigentlichen Vorlesungsinhalts in Form einer Diskussion zwischen dem Dozenten und den Teilnehmern möglich.

Materialbereitstellung für die Nachbereitungsphase

Zur Nachbereitung werden den Studierenden die in der Vorlesung präsentierten Inhalte inklusive der Tafelanschrift über die Lernplattform zur Verfügung gestellt. Die Inhalte können vom Dozenten noch nachträglich ergänzt oder kommentiert und ggf. auch korrigiert werden, ferner können themenverwandte Links bereitgestellt werden. Themen der Vorlesung können durch entsprechende Lehrpfade vertieft werden. Zur eigenen Lernkontrolle können von den Studierenden Tests bearbeitet werden. Vom Dozenten wird das Thema der nächsten Veranstaltung bekannt gegeben.

4 Aufbereitung des Textkorpus

4.1 Notwendigkeit der Aufbereitung des Textkorpus

Anhand der textlichen Inhalte des Lernnetzwerks „WiBA-Net“ soll in diesem Kapitel exemplarisch gezeigt werden, wie die darin enthaltenen Informationen zu einer Inhaltsbasis für die automatische Lehrpfadgenerierung aufbereitet werden können. Die Aufbereitung gliedert sich in die Konvertierung des Textkorpus, um eine einheitliche Datenbasis für die folgenden Berechnungen zu erhalten, und dessen anschließende textstatistische und linguistische Auswertung. Der Textkorpus bildet zusammen mit den so gewonnenen Informationen, die für das Auffinden und Weiterverarbeiten von Dokumenten genutzt werden können, die Inhaltsbasis.

Aus den bei der Aufbereitung gewonnenen Erkenntnissen lässt sich ableiten, wie neue Wissensseinheiten für eine zukünftige Systemerweiterung idealerweise aufgebaut sein sollten, um deren Einbeziehung in die vorhandene oder eine neue Inhaltsbasis zu vereinfachen. Ebenfalls lässt sich definieren, wie bereits anderweitig vorhandene Wissensseinheiten für die Nutzung mit dem vorgeschlagenen System ausgelesen und konvertiert werden können. Die einzelnen Schritte der Aufbereitung können automatisch ablaufen, jedoch muss der für das eigentliche Auslesen der Inhalte zuständige Programmteil an das jeweilige Format der vorliegenden Texte angepasst werden.

Die aus den Inhalten des WiBA-Net extrahierten Texte bzw. die aus ihnen gewonnenen Wortformen werden im Rahmen der automatischen Lehrpfadgenerierung

- bei der Bestimmung der Signifikanz von Kookkurenzen,
- bei der Bestimmung von Ähnlichkeiten zwischen Begriffen,
- bei der Suchterweiterung,
- bei der Booleschen Suche,
- bei der Bestimmung von Ähnlichkeiten zwischen Wissensseinheiten und
- bei der Strukturierung der erzeugten Lehrpfade

eingesetzt.

Die Konvertierung der Wissensseinheiten und die Durchführung der grundlegenden textstatistischen Berechnungen erfolgt in mehreren Schritten. Eine Übersicht der Schritte und der in jedem Schritt erhaltenen Zwischenergebnisse findet sich in Tabelle 4.1.

Ziele der Konvertierung der Menge der Wissensseinheiten sind die Vereinheitlichung bzw. Normierung verschiedener Schreibweisen, die Zerlegung des Textes in Sätze, die Bestimmung von Wortstämmen und die Zerlegung von Komposita. Außerdem werden für die einzelnen Wissensseinheiten Dokumentvektoren der enthaltenen Begriffe erzeugt und gespeichert. An die Konvertierung

schließt sich die textstatistische Auswertung der erhaltenen Sätze und Begriffe an. Hierbei werden zunächst die Signifikanzwerte der gefundenen Kookkurenzen von Wortpaaren berechnet. Für jeweils zwei Begriffe kann auf der Basis ihrer Kookkurenzen dann ein Ähnlichkeitswert bestimmt werden.

Tabelle 4.1: Schritte und Zwischenergebnisse bei der Content-Extraktion aus Wissensseinheiten

Ausgangsmaterial: Wissensseinheiten als HTML-Dateien	
Schritt	Ergebnis
1. Auslesen der Inhalte	Trennung Inhalt – Format
2. Normierung	normierte Sätze inkl. jeweils Dateiname (I)* Liste aller enthaltenen Worte (II)
3. Stemming	Liste aller gefundenen Stammformen (III) Zuordnung Ableitung → Stammform (IV) Liste aller Worte, die weder Ableitungen noch Stammformen haben (V) Datei mit normierten Sätzen → Ersetzen aller Worte durch die Stammformen (Ia) Zuordnung Wort → Dateien, die das Wort enthalten (inverse Liste) (VI) Liste aller Worte in Stammform (IIa)
4. Entfernen von Ausschlussworten	Häufigkeiten der Worte (VII) Ausschlussworte (100 häufigste Worte) (VIIa)
5. Kompositazerlegung	Zuordnung Komposita → Konstituenten (VIII)
6. Bestimmung der Kookkurenzen	Liste der Satzkookkurenzen (IX) Liste der Nachbarkookkurenzen (X) Vereinigung der Kookkurenzlisten (IX/Xa)
7. Bestimmung der Ähnlichkeiten jeweils zweier Begriffe	Ähnlichkeitswerte für Begriffspaare (IX/Xb)
8. Bestimmung der Dokumentvektoren	Vektordateien (XI)
Ergebnis: Datenbasis für die automatische Generierung von Lerneinheiten	

* Römische Ziffern weisen auf gespeicherte Daten hin, die später weiterverwendet werden. Kleinbuchstaben bezeichnen die Ableitung einer Datei aus einer anderen.

4.2 Konvertierung der Wissensseinheiten

4.2.1 Auslesen der Inhalte

Bei der Erstellung der WiBA-Net-Wissenseinheiten erfolgte keine Trennung zwischen Inhalt und Layout²⁶. Die WiBA-Net-Inhalte liegen als HTML-Dateien vor²⁷. Diese weisen bei der Anzeige im Webbrowser zwar eine einheitliche Formatierung auf, sind jedoch in ihrem inneren Aufbau unterschiedlich organisiert. Zur Erstellung der Wissensseinheiten des WiBA-Net wurden Formatvorlagen für unterschiedliche Seitentypen, die eine bedarfsgerechte Platzierung von Texten und Bildelementen auf den Seiten ermöglichen, zur Verfügung gestellt. Diese Vorlagen wurden zum Teil von den einzelnen Autoren noch an individuelle Erfordernisse angepasst²⁸. Der unterschiedliche Aufbau der Dateien muss beim Auslesen der Inhalte berücksichtigt werden: Inhalte werden nicht an fixen Positionen ausgelesen, sondern müssen zunächst überhaupt als Inhalte identifiziert werden.

In HTML-Dateien wird zwischen den eigentlichen Inhalten und strukturierenden oder formatierenden Elementen, den *Tags*, unterschieden. HTML-Tags bestehen aus einem in spitze Klammern eingeschlossenen Bezeichner, z.B. „<p>“, „<div>“ oder „<table>“, der die Eigenschaften des vom Tag ausgezeichneten Inhalts festlegt. Grundprinzip bei der Auszeichnung eines Textes in HTML ist, dass alle Inhalte zwischen einem öffnenden und einem schließenden Tag stehen. So wird in dem Ausdruck <p>Dies ist ein Absatz.</p> dem Inhalt durch das <p> („*paragraph*“) ein neuer Absatz zugewiesen. Dieser wird mit dem schließenden Tag </p> – zu erkennen am Schrägstrich – wieder beendet.

Für die vorliegende Arbeit ist es – mit Ausnahme der Überschriften – unerheblich, welcher Art der ausgelesene Text ist: Fließtext, Text aus Tabellen oder Bildunterschriften²⁹. Aus diesem Grund kann der Text durch sequenzielles Abarbeiten des HTML-Quellcodes extrahiert werden. Dabei werden alle in spitze Klammern („<“, „>“) eingeschlossenen Bereiche als HTML-Tags, also Formatierung behandelt, Bereiche außerhalb dieser Klammern stellen den Inhalt dar. Da vereinzelt im Inhaltsbereich spitze Klammern als „größer als“- oder „kleiner als“-Zeichen auftreten, werden diese Zeichen bei der Extraktion gesondert berücksichtigt, indem geprüft wird, ob jeweils ein Paar der

26 Mit Ausnahme der Inhalte, die mit dem an der Universität Stuttgart entwickelten System LORe erstellt wurden [BORRMANN U. SCHWARTE 2003]. Dieses System generiert die HTML-Seiten aus einer Datenbank. Im Rahmen der vorliegenden Arbeit wurde allerdings auch für diese Inhalte die HTML-Version der jeweiligen Wissensseinheit ausgewertet.

27 Der HTML-Quellcode einer exemplarischen WiBA-Net-Wissenseinheit sowie ein Screenshot von deren Anzeige im Webbrowser finden sich in Anhang H.

28 Als Beispiel sei der Seitentitel genannt, wie er im HTML-Tag „<title>“ festgehalten wird. Ursprünglich war hier für alle Seiten die Formulierung „WiBA-Net – Werkstoffe im Bauwesen und in der Architektur“ vorgesehen. Es finden sich jedoch mindestens vier verschiedene Versionen dieser Formulierung.

29 Eine eindeutige Erkennung der Textart ist – wiederum mit Ausnahme der Überschriften – auch oft nicht möglich.

Klammern vorhanden ist. Die zwischen den HTML-Tags stehenden Inhalte bestehen in der Regel aus Fließtext, der häufig mehrere Sätze enthält. Die zwischen zwei Tags ausgelesenen Inhalte werden jeweils als ein neuer Absatz betrachtet und in einer für alle Wissensseinheiten gemeinsamen Datei gespeichert, wobei ein Absatz jeweils auf eine Zeile geschrieben wird. Den Absätzen einer Wissensseinheit wird zur späteren Zuordnung jeweils der Dateiname der Wissensseinheit vorangestellt. Dabei ist zu beachten, dass auch an Stellen, die keinen neuen Absatz darstellen, fälschlich neue Unterabschnitte erkannt werden können. Dies geschieht in der Regel bei Tags, die nur der Formatierung (z.B. fett oder kursiv, Hoch- oder Tiefstellen, Änderung der Schriftgröße) des umgebenen Textes dienen. Zum Beispiel würde der Text

Dieser Satz enthält ein **fett** gedrucktes Wort.

der als HTML-Code wie folgt aussehen könnte:

```
<p>Dieser Satz enthält ein <b>fett</b> gedrucktes Wort.</p>
```

bei einer Text-Extraktion wie oben beschrieben folgende Satzfragmente liefern:

- Dieser Satz enthält ein
- fett
- gedrucktes Wort.

Eine nachträgliche Zuordnung dieser Fragmente zu einem Absatz bzw. einem Satz ist nicht mehr möglich, da auch eine andere, neue Sinnabschnitte definierende Form der Strukturierung – z.B. eine Tabelle – vorgelegen haben könnte. Deswegen muss der Text so vorformatiert werden, dass bestimmte, ausschließlich formatierende Tags weitgehend eliminiert werden, bevor die eigentliche Extraktion beginnt. In manche der Dateien ist zudem ein standardisierter Javascript-Programmcode eingebettet. Dieser wird ebenfalls vor der Weiterverarbeitung entfernt, da er zwischen umschließenden Tags steht und andernfalls zum Inhalt gezählt werden würde.

Die erste Zeile des extrahierten Inhalts entspricht dem Inhalt des sog. Title-Tags, des Titels der HTML-Seite, und enthält nur den Namen des Lernnetzwerks WiBA-Net. Die Zeile wird deswegen nicht berücksichtigt. Die nächste Zeile entspricht der Überschrift der Inhaltsseite. Sie wird zusätzlich gesondert festgehalten. Bei der späteren Lehrpfadgenerierung bekommen Seiten, die aufgrund eines in ihrer Überschrift vorkommenden Suchbegriffs gefunden wurden, eine höhere Priorität. Die übrigen Zeilen stellen die eigentlichen Inhalte der Wissensseinheit dar. Bei vielen Dateien befindet sich am Ende der Seite noch eine Zeile mit einem Copyright-Vermerk. Dieser wird, wenn vorhanden, ebenfalls entfernt.

4.2.2 Normierung der Inhalte

Als nächster Schritt folgt nun das Normieren der Inhalte auf eine einheitliche Form. Das Ziel ist eine Reduzierung der verwendeten Zeichen und eine einheitliche Schreibweise, sodass dieselben Worte unabhängig von ihrer Notation vom System als gleich erkannt werden. Es geht dabei in erster Linie um die Groß- und Kleinschreibung sowie die Notation von Sonderzeichen. Eine Rechtschreibprüfung findet nicht statt.

Im Folgenden wird die Groß- und Kleinschreibung nicht weiter beachtet³⁰, deswegen wird der Text zunächst in Kleinbuchstaben umgewandelt. Ziffern werden aus dem Text entfernt. Dies ist notwendig, da Zahlenwerte quasi zufällig in den Inhalten verteilt sind. Es bestehen folglich keine inhaltlichen Zusammenhänge mit Worten aus dem Text. Eine Ausnahme stellt jedoch die Bezeichnung von Normen (z.B. „DIN 1045“) oder Baustoffen (z.B. „C 30/37“) dar. Phrasen, also Zusammensetzungen aus mehreren Worten wie „DIN 1045“, werden in der vorliegenden Arbeit nur als Einzelworte berücksichtigt. Eine Einbeziehung dieser Zahlen in die textstatistische Auswertung wäre deswegen nicht möglich, denn die Zahlen könnten auf diese Weise nicht eindeutig von anderen vorkommenden Zahlen unterschieden werden.

Viele Sonderzeichen können über Tastaturen nicht direkt eingegeben werden, insbesondere wenn das Tastaturlayout nicht der verwendeten Sprache entspricht. So kann z.B. das „Vébé-Maß“ zwar mit einer deutschen Tastatur eingegeben werden, die dänische Schreibweise der im WiBA-Net erwähnten „Øresund-Brücke“ kann jedoch nur softwareseitig durch Einfügen eines Sonderzeichens realisiert werden. Um die Eingabe der Sonderzeichen und ihre einheitliche Darstellung im Webbrowser zu vereinfachen, werden Sonderzeichen in HTML durch so genannte *Entities* ersetzt. Es gibt hierbei zwei Möglichkeiten: Entweder wird die Unicode-Nummer des Zeichens verwendet oder ein so genanntes „*named entity*“, das für häufig verwendete Zeichen definiert ist [SELFHTML 2007].

Der deutsche Umlaut „ä“ wird als *named entity* zu „ä“, das erwähnte dänische „Ø“ zu „Ø“. Als Entity nach der Unicode-Nummer wird „ä“ zu „ä“. Die Ersetzung der Sonderzeichen wird von HTML-Editoren meist automatisch vorgenommen. Zählt man die nicht codierte Form³¹ der direkt eingebbaren Sonderzeichen mit, so erhöht sich die Zahl der Notations-Möglichkeiten noch. Es gibt z.B. drei Möglichkeiten, ein umgelautetes „a“ in HTML darzustellen: „ä“, „ä“ und „ä“. Da alle Schreibweisen gleichwertig sind, sollten sie bei einem Vergleich,

30 In der Literatur werden bei bestimmten sprachstatistischen Anwendungen ausschließlich oder vornehmlich Substantive betrachtet (vgl. z.B. [HEYER ET AL. 2006]). Da in der Terminologie des Bauingenieurwesens viele Fachausdrücke auch Verben und Adjektive sind, wird auf deren Berücksichtigung im Rahmen der vorliegenden Arbeit nicht verzichtet. Es entfällt dadurch ebenfalls die Berücksichtigung der Groß- und Kleinschreibung als Kriterium für die Identifikation von Substantiven.

31 Die nicht codierte Form wird von modernen Webbrowsern auch korrekt dargestellt.

z.B. im Rahmen einer Suche, vom vergleichenden Programm auch als gleich erkannt werden. Dies wird für die Inhalte des WiBA-Net erreicht, indem alle Sonderzeichen durch ihre Umschreibungen in den Buchstaben des ASCII-Zeichensatzes³² ersetzt werden. Ein umgelautetes „a“ wird also beispielsweise immer als „ae“ notiert, das „Vébé-Maß“ wird als „vebe-mass“ gespeichert. Auf diese Weise kann auch leicht der Zusammenhang zwischen einem umgelauteten Vokal und dessen Umlaut berücksichtigt werden, z.B. in den Worten „Stahl“ und „Stähle“. Dies ist bei der Bestimmung der Wortstämme wichtig (vgl. Abschnitt 4.2.3).

Danach werden häufige, eindeutige Abkürzungen aufgelöst und durch ihre Ausschreibung ersetzt. Die dadurch ermöglichte Worterkennung ist eher nebensächlich, da die abgekürzten Worte in der Regel sehr häufig auftreten und deswegen nur einen geringen Beitrag zur Erkennung nutzbarer Beziehungen in der statistischen Textauswertung liefern. Man entfernt auf diese Weise aber auch die in den Abkürzungen verwendeten Punkte. Diese können bei der folgenden Zerlegung des Textes in einzelne Sätze nicht mehr fälschlich für ein Satzende gehalten werden. Auf diese Weise erhöht sich die Wahrscheinlichkeit, dass ein Punkt tatsächlich ein Satzende markiert und nicht wegen einer anderen Bedeutung vorhanden ist. Frage- und Ausrufezeichen, die ebenfalls ein Satzende markieren, werden durch Punkte ersetzt, um die folgende Aufteilung in Sätze zu erleichtern. Aus dem Text werden abschließend noch alle anderen Interpunktionszeichen sowie verbleibende Sonderzeichen (z.B. „§“, „&“) entfernt. Der Text besteht am Ende der Normierung also nur noch aus Kleinbuchstaben nach dem ASCII-Zeichensatz, Leerzeichen und Punkten.

Allgemein ist bei der oben erläuterten Textbereinigung die Reihenfolge wichtig. Entfernt man z.B. zuerst alle Kaufmanns-Und-Zeichen („&“) aus dem Text, lassen sich in späteren Schritten die HTML-Sonderzeichen („Entities“), die grundsätzlich mit „&“ beginnen (z.B. „ö“, „ä“) nicht mehr leicht identifizieren.

Abschließend wird der wie beschrieben normierte Text an den verbleibenden Punkten, die jetzt alle als Markierung eines Satzendes betrachtet werden können, aufgeteilt. Die so erhaltenen Sätze werden zur Weiterverarbeitung als Zeilen in eine Datei geschrieben, jede Zeile erhält dabei einen Vermerk dazu, aus welcher Datei sie stammt. Außerdem kann aus den normierten Texten eine Liste aller enthaltenen Worte erstellt werden. Diese Liste dient als Grundlage für das im folgenden Abschnitt beschriebene Stemming.

³² Da ja alle Groß- in Kleinbuchstaben umgewandelt wurden, sind die relevanten Buchstaben des ASCII-Zeichensatzes „abcdefghijklmnopqrstuvwxyz“.

4.2.3 Rückführung von Worten auf ihre Stämme

Beim Stemming werden die in der Liste aller Worte enthaltenen Ausdrücke nach Möglichkeit auf gemeinsame Grundformen zurückgeführt. Dabei wird nicht unbedingt die tatsächliche grammatische Stammform („Wurzel“ oder „Wortstamm“) bestimmt, sondern die aus den vorhandenen Worten kürzeste ableitbare Wortform. Für den geforderten Zweck, nämlich die Bestimmung von Ähnlichkeiten zwischen verschiedenen Dokumenten bzw. Dateien, ist diese Vorgehensweise vollkommen ausreichend. Es ist nicht entscheidend, ob ein Wort den eigentlichen Stamm darstellt, vielmehr ist es wichtig, das zwei abgeleitete Begriffe als zum selben Stamm gehörig und damit als semantisch identisch interpretiert werden können. Auf diese Weise muss auch keine manuell zu pflegende und zu ergänzende Datenbank mit tatsächlichen Stammformen geführt werden. Bei einer Erweiterung des Systems können neue Stammformen bzw. kürzeste Wortformen automatisch bestimmt werden. Das Verfahren zur Bestimmung der Stammformen ist in Abbildung 4.1 schematisch dargestellt.

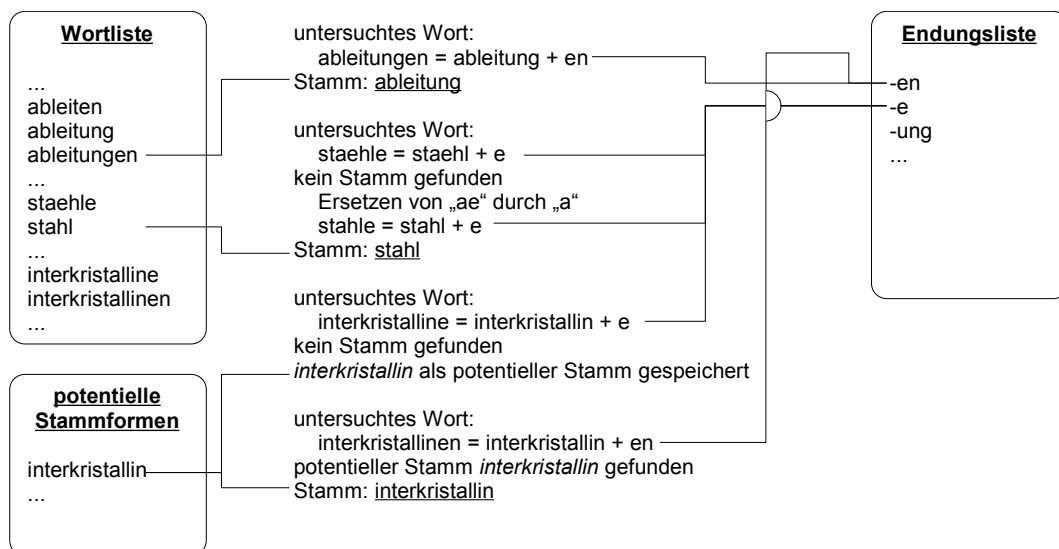


Abbildung 4.1: Schematische Darstellung des Algorithmus zur Bestimmung von Stammformen

Zunächst wird bei allen Worten untersucht, ob sie auf eine Flexions- oder Derivationsendung, also eine bestimmte Buchstabenkombination („Suffix“)³³, enden. Ist das der Falle, wird überprüft, ob es ein kürzeres Wort in derselben Liste gibt, das dem untersuchten Wort abzüglich der Flexionsendung entspricht. Dieses wird im Erfolgsfall als Stamm des untersuchten Wortes vermerkt. Wird auf diese Weise kein Stamm gefunden, werden ggf. enthaltene Umlaut-Umschreibungen („ae“, „oe“, „ue“) in den jeweils umgelauteten Buchstaben umgewandelt. Das Wort wird dann nochmals mit der

33 vgl. Anhang C

Liste verglichen. Eine eventuelle Übereinstimmung wird ebenfalls als Stamm vermerkt. Auf diese Weise findet man z.B. „stahl“ als Stamm für „staehle“.

Bei den Worten, bei denen das Abtrennen einer Endung und die Ersetzung der Umlaute nicht zum Finden einer vorhandenen Stammform führt, kann eine andere Methode zur Bestimmung der Stammform angewendet werden. Es wird für diese Worte überprüft, ob es ein Wort in der Liste gibt, von dem man eine andere Endung abtrennen kann und dabei dieselbe Kurzform erhält. Ein solches Wort wird dann als Stammform für beide Ableitungen gespeichert. Auf diese Weise erhält man z.B. aus den Ableitungen „interkristalline“ (mit der Endung „e“) und „interkristallinen“ (mit der Endung „en“) die Stammform „interkristallin“. Diese Wortform war vorher selbst nicht in der Wortliste enthalten.

Man erhält so eine Liste aller Stammformen mit ihren Ableitungen. Abschließend wird für die ermittelten Stammformen auf dieser Liste überprüft, ob diese selber an anderer Stelle wieder einer, noch kürzeren, Stammform zugeordnet wurden. Falls ja, werden die entsprechenden Listeneinträge zusammengefasst. Jeder Ableitung wird so nur die insgesamt kürzeste Stammform zugeordnet. Beispielsweise wird dem Wort „ableitungen“ der Stamm „ableitung“ zugeordnet. Zu diesem Wort wird aber, unter Berücksichtigung des Ausdrucks „ableiten“, der Stamm „ableit“ gefunden. Dieser stellt dann auch die Stammform von „ableitungen“ dar.

Zur Weiterverarbeitung werden drei Wortlisten gespeichert. Die erste Liste weist jeder Ableitung ihre Stammform zu. Die zweite enthält alle Stammformen, zu denen Ableitungen existieren. Die dritte enthält zur Kontrolle alle Worte, zu denen weder Ableitungen noch Stammformen gefunden werden konnten.

Auf dieser Basis wird jetzt die Datei, die die normierten Sätze enthält, umgeformt. Dabei werden die enthaltenen Worte durch ihre jeweiligen Stammformen³⁴ ersetzt. Zusätzlich wird eine inverse Datei gespeichert, die zu jeder Stammform jeweils die Dateinamen enthält, in denen das Wort vorkommt. Da die Gesamtmenge der Dateien mit 2040 relativ klein ist, wird auf eine Zerlegung der inversen Datei zur Zugriffsbeschleunigung verzichtet (vgl. Abschnitt 2.2.2.1). Für die weitere Bearbeitung wird ferner eine Liste aller Worte in gestemmter Form benötigt, die ebenfalls jetzt gespeichert werden kann. Diese Liste enthält für alle Ableitungen nur noch einmal die gemeinsame Stammform. Deshalb ist diese Liste wesentlich kürzer als die ursprüngliche Liste aller Wortformen³⁵.

Für die spätere statistische Auswertung werden außerdem die Häufigkeiten der Wortformen benötigt. Diese werden für jede der Stammformen ermittelt und gesichert. Da sehr häufige Worte (wie

³⁴ Im Folgenden wird durch den Begriff Stammform sowohl die Grundform einer abgeleiteten Wortform als auch die Wortform selbst bezeichnet, wenn sie identisch mit der Stammform ist.

³⁵ In der Liste mit den ungestemten Wortformen finden sich 28.053 Wortformen, in der Liste mit den Stammformen nur 19.347.

z.B. Artikel oder Konjunktionen) nur einen geringen Nutzwert als Suchbegriff haben, werden die einhundert häufigsten Worte („Ausschlussworte“³⁶) aus den Satzlisten entfernt. Es ist jedoch zu beachten, dass aufgrund des fachsprachlichen Charakters der WiBA-Net-Inhalte und des im Vergleich zu in der Literatur verwendeten Textkorpora (vgl. z.B. [ACKERMANN 2000]) geringen Umfangs des WiBA-Net-Textkorpus auch einige wichtige fachsprachliche Ausdrücke unter den einhundert häufigsten Begriffen sind (z.B. „Stahl“, „Beton“ oder „Holz“, vgl. Anhang D). Diese Begriffe sollen den Nutzern des Systems auch als Suchbegriffe zur Verfügung stehen, deshalb werden diese Begriffe beim Ausschließen der häufigsten Begriffe übersprungen.

4.2.4 Kompositazerlegung

Nachdem das Stemming abgeschlossen ist, können durch eine Kompositazerlegung die Bedeutungen einzelner Wortbestandteile erschlossen werden. Das Grundprinzip entspricht dem des Stemming: Es werden gefundene Wortformen mit einer Liste existierender Wortformen verglichen. Findet sich eine Übereinstimmung, wird die gefundene Form als richtig akzeptiert. Im Unterschied zum Stemming müssen bei der Kompositazerlegung neben den Endungen der Worte auch die so genannten Fugenelemente [MEIBAUER ET AL. 2002] berücksichtigt werden. Diese werden zwischen zwei Konstituenten eines Kompositums eingefügt, i.d.R. um die Aussprache zu erleichtern. Ein Beispiel für ein solches Fugenelement ist das „s“ zwischen „Entmischung“ und „Gefahr“ in „Entmischungsgefahr“.

Für die Kompositazerlegung werden nacheinander die einzelnen Begriffe der gestemmtten Wortliste abgearbeitet. Für jeden Begriff bzw. Wortstamm in der Liste wird geprüft, ob er in einem der anderen Worte der Liste enthalten ist. Worte, bei denen das der Fall ist, werden daraufhin überprüft, ob sie ein Kompositum sind, das einen Konstituenten enthält, der dem gerade betrachteten Begriff entspricht. Zu beachten ist, dass zu sehr vielen Komposita nur eine Wortform im Textkorpus vorhanden ist, also beim in Abschnitt 4.2.3 erläuterten Stemming keine Stammform gefunden werden konnte. Diese Worte verfügen also, im Gegensatz zu den gestemmtten Begriffen, noch über Flexionsendungen. Beim Test auf eine Übereinstimmung muss also das Ende des Wortes zusätzlich daraufhin überprüft werden, ob es in den definierten Endungen vorkommt. Hat man durch Entfernen des betrachteten Begriffs und der ggf. vorhandenen Endung einen Wortrest gefunden, wird wie beim Stemming überprüft, ob dieser Wortrest, ggf. ohne Endung bzw. Fugenelement, in der Wortliste vorkommt. Kann für ein Wort auf diese Weise festgestellt werden, dass es nur aus Wortstämmen sowie Fugenelementen und Endungen besteht, so wird es als Kompositum berücksichtigt. Für alle gefundenen Komposita werden die enthaltenen Konstituenten gespeichert.

³⁶ in der Literatur auch als „Stoppworte“ bezeichnet

Die Vorgehensweise bei der Kompositazerlegung wird im Folgenden an einem Beispiel erläutert, das in Abbildung 4.2 schematisch dargestellt ist.

Bei der Abarbeitung der Liste der Stammformen wird der Begriff „Technik“ betrachtet. Dieser findet sich u.a. in dem in derselben Liste enthaltenen Wort „Herstellungstechniken“ wieder. Es wird deswegen untersucht, ob das Wort „Herstellungstechniken“ ein Kompositum ist. Das Wort wird zunächst an der Stelle des Begriffs „Technik“ aufgeteilt. Man erhält die Elemente „Herstellungs-“ und „-en“. Da es sich bei „-en“ um ein Element der Endungsliste handelt, braucht nur der erste Wortrest „Herstellungs-“ weiter betrachtet zu werden. Dieses lässt sich in „Herstellung“ und das Fugenelement „s“ aufteilen. Es muss also nur der Wortrest „Herstellung“ weiter betrachtet werden. Für diesen findet sich die Stammform „herstell“. Damit sind alle Bestandteile des Wortes „Herstellungstechniken“ zugeordnet. Es handelt sich um ein Kompositum aus den Konstituenten „herstell“ und „technik“. Die Informationen über das Fugenelement und die Flexionsendung sind für den weiteren Verlauf irrelevant.

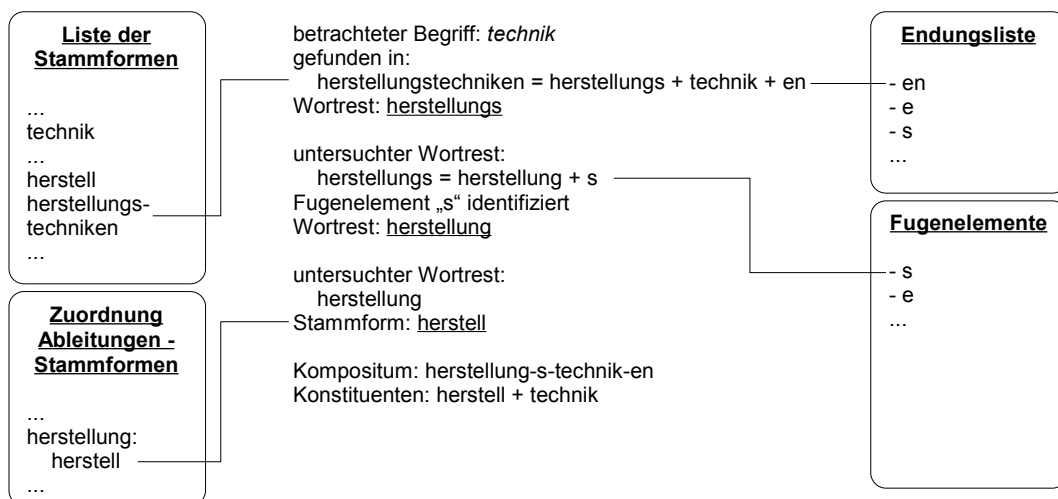


Abbildung 4.2: Schematische Darstellung einer Kompositazerlegung

Ein Versuch, analog zum Stemming unbekannte Wortformen als Konstituenten zu identifizieren, wenn sie in mehreren Komposita vorkommen, scheiterte. Bei dieser Vorgehensweise wurden sehr viele nicht existierende Worte als angebliche Konstituenten identifiziert. Aus diesem Grund bleiben einige Komposita unzerlegt. Dieses Problem kann potentiell durch eine Erweiterung der Wortliste gelöst werden. Diese tritt ein, wenn neue Inhalte mit noch nicht in der Inhaltsbasis vorhandenen Worten in das System eingepflegt werden.

4.3 Textstatistische Berechnungen

4.3.1 Bestimmung der Kookkurenzen

Untersucht werden die Satz- und Nachbarkookkurenzen aller nach dem Entfernen der Ausschluss- worte im Textkorpus verbleibenden Worte. Bei den Satzkookkurenzen werden die Konstituenten der im Satz enthaltenen Komposita mit berücksichtigt. Vor der Bestimmung der Anzahl des gemeinsamen Auftretens der Wortpaare werden alle Komposita durch ihre Konstituenten ergänzt. Dies führt zwar dazu, dass die Konstituenten und das Kompositum als häufig gemeinsam auftretend erkannt werden. Dies ist ja aber auch völlig richtig und führt damit nicht zu einer Verfälschung der Ergebnisse. Darüber hinaus erhält man jedoch auch Kookkurenzen – und damit ggf. Verknüpfungen – der anderen Worte eines Satzes mit den Konstituenten. So kann z.B. durch den Satz „Eichenholz ist eine Holzart mit hoher Festigkeit und mäßigem Schwindverhalten.“ durch das gemeinsame Auftreten der Konstituenten „Holz“ und „Schwind(en)“ die Signifikanz dieser Kookkurenz bzw. die Verknüpfung zwischen diesen beiden Begriffen erhöht werden.

Für die Berechnung nach Formel (2.6) bzw. Formel (2.15) müssen die Häufigkeiten des gemeinsamen Auftretens der zu betrachtenden Wortpaare bestimmt werden. Dies ist einmal für direkt benachbarte Terme durchzuführen und einmal für die Terme aller Sätze. Neben der Anzahl des gemeinsamen Auftretens der beiden jeweils untersuchten Begriffe wird auch die jeweilige Anzahl an Sätzen bzw. Paarungen benachbarter Worte benötigt, die nur den ersten bzw. nur den zweiten untersuchten Term enthalten. Bei der Ermittlung der Anzahl muss bei den Satzkookkurenzen beachtet werden, dass die Anzahl sich gegenüber den gespeicherten Satzlisten geändert hat, da die Sätze für die aktuelle Berechnung um die Konstituenten der enthaltenen Komposita erweitert wurden.

Für die Nachbarkookkurenzen werden immer zwei benachbarte Worte bis zum Ende eines Satzes betrachtet. Bei einem Satz mit n Worten entstehen so $n-1$ Kombinationen. Bei der Bestimmung der Satzkookkurenzen werden alle unterschiedlichen Worte in einem Satz betrachtet und das gemeinsame Auftreten mit allen anderen Worten dieses Satzes gezählt. Ein Satz mit n Worten liefert somit

$$\binom{n}{2} = \frac{(n^2 - n)}{2} \quad (4.1)$$

Kombinationen. Da die Reihenfolge der beiden betrachteten Begriffe irrelevant ist³⁷, werden zur Reduktion der zu berechnenden Kombinationen die beiden Begriffe alphabetisch sortiert und für diese Kombination dann die Häufigkeiten des ersten und des zweiten Terms gespeichert und die

³⁷ Auch bei den Nachbarkookkurenzen wird nicht zwischen linken und rechten Nachbarn unterschieden.

Anzahl der gemeinsamen Auftreten hochgezählt. Die Listen der Paarungen und ihrer jeweils drei Häufigkeitswerte werden für Satz- und Nachbarkookkurenzen separat zwischengespeichert.

Im nächsten Schritt werden für die Paarungen jeweils die Signifikanzen nach Formel (2.6) bzw. Formel (2.15) berechnet. Für jeden Term werden alle Kookkurenzen nach ihrem Signifikanzwert geordnet gespeichert. Die berechneten Signifikanzen beruhen auf der jeweiligen Wahrscheinlichkeit, dass die beiden betrachteten Begriffe, die k-mal zusammen auftreten, genau diese k-mal zusammen auftreten. Aus diesem Grund kann kein bestimmter Signifikanzwert angegeben werden, ab dem zwei Begriffe als verwandt betrachtet werden können. Folglich müssten für die nachfolgenden Berechnungen zunächst alle errechneten Signifikanzwerte zur Verfügung stehen. In der Praxis treten jedoch, insbesondere bei sehr häufigen Begriffen, hohe Anzahlen von Kookkurenzen auf. Dieser Effekt wird durch die eingefügten Konstituenten der Komposita noch erhöht³⁸. Um die Rechenzeiten nicht unnötig zu erhöhen, werden deshalb nur die jeweils 250 Satz- bzw. Nachbarkookkurenzen mit den höchsten Signifikanzwerten für die weitere Verarbeitung gespeichert.

4.3.2 Vereinigung von Satz- und Nachbarkookkurenzen

Wie in Abschnitt 4.3.1 beschrieben, werden für die vorhandenen Wissenseinheiten zwei verschiedene Arten von Kookkurenzen berechnet. Diese Kookkurenzen beschreiben in der Regel unterschiedliche Relationen zwischen den jeweils betrachteten Begriffen. So kann es sich bei Satzkookkurenzen z.B. eher um Synonyme handeln, bei Kookkurenzen benachbarter Worte eher um Begriffe und deren Eigenschaften [HEYER ET AL. 2006]. Im vorliegenden Anwendungsfall werden gerade diese Informationen aber nicht für die Verarbeitung der Wissenseinheiten genutzt.

Trotzdem ist eine Berechnung sowohl der Satz- als auch der Nachbarkookkurenzen sinnvoll. Bei der Bestimmung der Satzkookkurenzen können nämlich auch die Konstituenten der Komposita mit betrachtet werden. Bei Nachbarkookkurenzen ist dies nicht sinnvoll, da auf diese Weise hauptsächlich die ein Wort bildenden Konstituenten miteinander verglichen würden. Es entsteht bei beiden Kookkurentypen jedoch eine unterschiedliche Ordnung der verwandten Begriffe. Bei Satzkookkurenzen stehen wie oben beschrieben eher Synonyme vorne, bei Kookkurenzen benachbarter Worte eher Eigenschaften. Eine getrennte Berechnung der Kookkurenzen und deren folgende Vereinigung erzielt also die besten Treffer aus beiden Bereichen.

Berechnet man den Mittelwert aller Signifikanzwerte der zur Verfügung stehenden höchsten Werte der Satz- und Nachbarkookkurenzen, so erhält man für die Satzkookkurenzen einen Mittelwert von 0,472, für die Nachbarkookkurenzen einen Mittelwert von 0,610. Dieser Unterschied in der Größenordnung muss bei einer Vereinigung der Kookkurentypen berücksichtigt werden, wenn man

³⁸ Für den Begriff Stahl werden z.B. 4.324 Satzkookkurenzen gefunden.

Satz- und Nachbarkookkurrenzen gleichwertig einfließen lassen will. Die diesbezügliche Vorgehensweise wird im Folgenden erläutert. Es tritt jedoch auch ein umgekehrter Effekt auf, der in Abbildung 4.3 dargestellt ist.

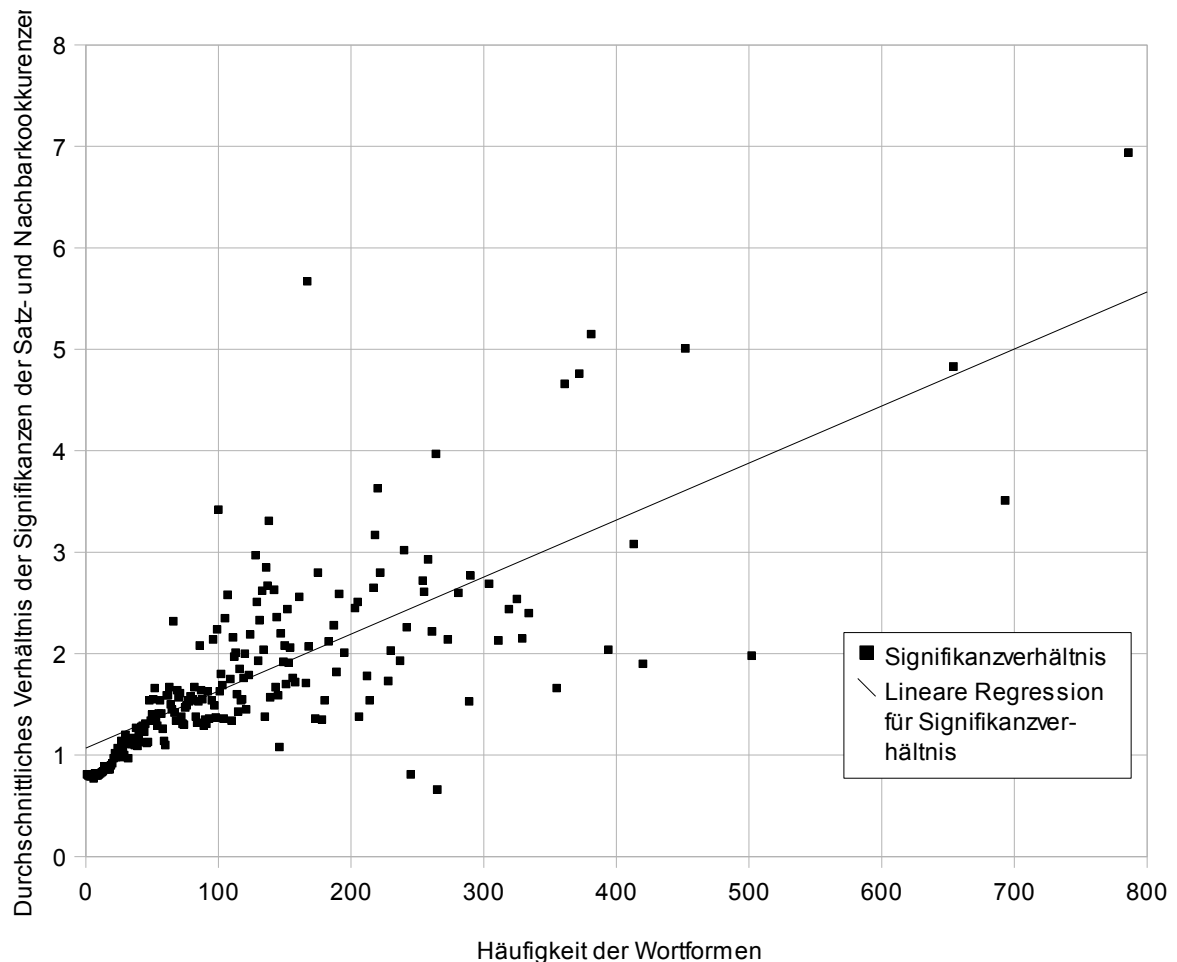


Abbildung 4.3: Verhältnis der durchschnittlichen Signifikanzwerte der Satz- zu denen der Nachbarkookkurrenzen, in Abhängigkeit von der Häufigkeit des Grundbegriffs

In diesem Diagramm ist das Verhältnis der Signifikanz der Satz- zu denen der Nachbarkookkurrenzen in Abhängigkeit von der Häufigkeit des Ausgangsbegriffs dargestellt. Man erkennt eine Verdichtung der Werte im unteren und eine starke Streuung im oberen Bereich. Eine steigende Tendenz ist eindeutig zu erkennen. Dies wird durch die eingefügte Regressionsgerade veranschaulicht. Die stark streuenden Werte im oberen Bereich sind darauf zurückzuführen, dass höhere Häufigkeitswerte nur wenige zugehörige Begriffe aufweisen und Schwankungen für eine Häufigkeit nicht aufgefangen werden können. So kommt der zweitgrößte Häufigkeitswert (693 Vorkommen)

nur bei dem Begriff „Stahl“ vor. Demgegenüber gibt es 2993 Begriffe, die die Häufigkeit 2, den zweitniedrigsten Häufigkeitswert, aufweisen³⁹.

Der untere Kurvenabschnitt mit Signifikanzverhältnisswerten kleiner eins umfasst Häufigkeitswerte bis zu 23 Vorkommen⁴⁰. Worte, die 23 mal oder seltener vorkommen, machen jedoch 94,15 % des Textkorpus des WiBA-Net aus. Dies erklärt, wieso der o.a. durchschnittliche Signifikanzwert für die Nachbarkookkurenzen von 0,610 größer ist als der für die Satzkookkurenzen von 0,472, obwohl nach Abbildung 4.3 augenscheinlich fast alle Verhältnisswerte größer als eins sind, also auf größere Satz- als Nachbarkookkurenzen hindeuten.

Um die beiden Signifikanztypen mit gleichem Gewicht zu vereinigen ist es deshalb angezeigt, die Satz- bzw. Nachbarkookkurenzen für jeweils einen Begriff auf dieselbe Größenordnung umzurechnen. Ansonsten werden die jeweils durchschnittlich kleineren Werte benachteiligt, wenn abschließend wieder die 250 höchsten gemeinsamen Signifikanzwerte gespeichert werden. Für jeden Begriff wird deshalb der Mittelwert der Signifikanzen der Satzkookkurenzen $MW_{sig(A,B),S}$ und der Nachbarkookkurenzen $MW_{sig(A,B),NB}$ berechnet. Dann werden die Signifikanzen der einzelnen Nachbarkookkurenzen $sig(A,B)_{NB}$ mit dem Quotienten aus dem Mittelwert der Signifikanzen der Satzkookkurenzen und dem Mittelwert der Signifikanzen der Nachbarkookkurenzen multipliziert. Es gilt:

$$sig(A,B)'_{NB} = sig(A,B)_{NB} \cdot \frac{MW_{sig(A,B),S}}{MW_{sig(A,B),NB}} \quad (4.2)$$

Anhand dieser an die Größenordnung der Satzkookkurenzen angepassten Nachbarkookkurenzen $sig(A,B)'_{NB}$ werden dann die 250 höchsten Signifikanzwerte beider Kookkurenztypen bestimmt und zur Weiterverarbeitung gespeichert.

4.3.3 Bestimmung der Ähnlichkeit von Begriffen

Die Kookkurenzen eines Begriffs stellen dessen Kontext dar (vgl. [HEYER ET AL. 2006]). Ein Begriff hat eine bestimmte Zugehörigkeit zu allen Begriffen seines Kontextes, die durch den Signifikanzwert der Kookkurenz der beiden Begriffe ausgedrückt wird. Man kann ferner als weiteres Maß der Zusammengehörigkeit zweier Begriffe die Ähnlichkeit ihrer Kontexte bestimmen (vgl. [ACKERMANN 2000]). Die Bestimmung eines solchen Ähnlichkeitsmaßes hat den Vorteil, dass die beiden Begriffe nicht wie zwei Kookkurenzen irgendwo im Textkorpus in demselben Satz vorkom-

³⁹ vgl. hierzu auch Abbildung 4.4

⁴⁰ Mit zwei Ausnahmen für die Häufigkeiten 245 und 265.

men müssen. Eine Ähnlichkeit kann demzufolge auch für Begriffe, die nicht in gemeinsamen Wissenseinheiten vorkommen, ermittelt werden, wenn die jeweiligen Kontexte verglichen werden.

Grundannahme dafür ist, dass zwei Worte, die eine ähnliche Bedeutung haben oder die ähnliche Begriffe bezeichnen, häufig mit denselben Worten in einem Satz vorkommen. Begriffe, die hohe Ähnlichkeiten aufweisen, können im weiteren Verlauf zur Suchterweiterung herangezogen werden.

Für die Begriffe der Wortliste werden die Ähnlichkeiten ihrer Kookkurenzen nach Formel (2.19) berechnet. Da eine Ähnlichkeit nach der Definition in HEYER ET AL. 2006 nur für Dokumente berechnet werden kann, werden die Kookkurenzen der beiden zu vergleichenden Terme als Dokumente betrachtet. Es wird für jeden Term seine relative Häufigkeit n_f in diesem Dokument bzw. Satz von Kookkurenzen berechnet, die Abhängig vom Umfang der Kookkurenzen ist. Ferner wird die Inverse Dokumentfrequenz (IDF) aus der Anzahl aller Wissenseinheiten und der Häufigkeit des betrachteten Terms bestimmt. Die Multiplikation von n_f und IDF ergibt nach Formel (2.18) das Gewicht des jeweiligen Terms. Die Ähnlichkeit der beiden Kookkurenzensätze wird dann aus den gewichteten Termen nach dem Kosinusmaß gemäß Formel (2.19) bestimmt.

Ein an dieser Stelle auftretendes Problem ist die Rechenzeit, die bei einem Vergleich eines Begriffs nach der obigen Beschreibung mit jeweils allen anderen Begriffen des Textkorpus benötigt wird. Aus diesem Grund wird im prototypischen Programmsystem die beschriebene Ähnlichkeitsbestimmung bei Bedarf bei der Eingabe einer Suchanfrage durchgeführt. Die einmal gefundenen Ergebnisse werden gespeichert und bei späteren Berechnungen wiederverwendet. Dies ist jedoch nur zulässig, solange der Textkorpus nicht geändert oder erweitert wird.

Analog zur Bestimmung der Ähnlichkeit zwischen Begriffen kann auch die Ähnlichkeit zwischen tatsächlichen Dokumenten, in diesem Fall den Wissenseinheiten, bestimmt werden. Für alle Wissenseinheiten des WiBA-Net werden die Ähnlichkeitswerte zu allen anderen Wissenseinheiten gespeichert. Diese Ähnlichkeiten lassen sich bei der Bestimmung weiterführender Seiten nutzen, indem Seiten, die einen bestimmten Ähnlichkeitswert zu einer anderen Seite überschreiten, dieser in einem automatisch generierten Lehrpfad als weiterführende Seite hinzugefügt werden.

4.4 Datenspeicherung

Damit für den Nutzer des Systems aus einer Suchanfrage ein Lehrpfad generiert werden kann, werden verschiedene Daten benötigt. Diese müssen auf dem Server in geeigneter Weise vorgehalten werden. Unterschieden wird zwischen den Inhalten und den aus den Inhalten gewonnenen textstatistischen Daten.

Nach der Normierung, dem Stemming und der Kompositazerlegung wird für jede Wissensseinheit ein Dokumentvektor gespeichert, der die um die Ausschlussworte bereinigten normierten und gestemmtten Worte sowie die Konstituenten der Komposita enthält. Da für die verschiedenen angewandten Verfahren zur Ähnlichkeitsmaßbestimmung entweder die Information, ob ein Term überhaupt in einem Dokument enthalten ist, oder aber die Häufigkeit der einzelnen Terme in jedem Dokument benötigt werden, werden in den Wissensseinheiten mehrfach vorkommende Terme in den Vektordateien auch mehrfach abgespeichert. Es wird außerdem festgehalten, welche Begriffe in dieser Vektordatei aus der Kompositazerlegung stammen, damit diese später bei bestimmten Berechnungen (durchschnittliche Häufigkeit von enthaltenen Begriffen) ausgeklammert werden können.

Das Programm zur automatischen Lehrpfadgenerierung greift nicht direkt auf die Wissensseinheiten zu, sondern nur auf die abgeleiteten Informationen. Dies sind

- ein Volltext-Index, der jedem Suchbegriff die Dateien zuweist, die den Suchbegriff enthalten,
- eine Vektordatei für jede Wissensseinheit, die nur die sinntragenden Terme dieser Datei enthält,
- die Häufigkeiten des Auftretens der einzelnen Terme im Gesamtkorpus,
- Wortstämme zu allen Termen,
- Konstituenten zu allen Komposita
- Kookkurenzen und deren Signifikanzen zu allen Termen und
- Ähnlichkeiten der Suchbegriffe mit anderen Begriffen.

Die eigentlichen Wissensseinheiten stehen im Zentrum der Lehrpfadgenerierung. Sie werden nur für die Anzeige im Webbrowser benötigt und können am einfachsten in Form von HTML-Seiten auf einem Webserver vorgehalten werden. Dem für die Ausgabe der Ergebnisse – also die Anzeige der Lehrpfade – zuständigen Programmteil muss die jeweilige URL zur Verfügung stehen, um die Seite dem Nutzer anzuzeigen.

Soll das Layout der Wissensseinheiten dynamisch während der Anforderung der Seiten generiert werden, so können die Inhalte der Wissensseinheiten auch in einem anderen, maschinenlesbaren Format, z.B. XML, vorgehalten werden oder alternativ in einer Datenbank gespeichert werden. Es wird dann noch eine Umformungsschicht zur Darstellung der Ergebnisse zwischen den Webserver und die gespeicherten Inhalte geschaltet. Diese kombiniert beim Aufruf einer benötigten Wissensseinheit den Inhalt aus der Datei oder der Datenbank mit einer Seitenvorlage zur Formatierung.

Damit vom System erstellte Lehrpfade mit dem entsprechenden Anzeigeprogramm betrachtet werden können und durch die Lehrpfade navigiert werden kann, müssen sie ebenfalls, wenn auch nur temporär, auf dem Server gespeichert werden. Dies erfolgt in Form von XML-Dateien, die die Struktur des Lehrpfads und die Speicherorte der enthaltenen Dateien umfassen. Sie beinhalten ferner die Überschrift der jeweiligen Seite, um diese im Navigationsbereich anzeigen zu können.

Automatisch generierte Lehrpfade können anstatt von Lernern auch von Fachexperten (z.B. Dozenten) erstellt und dann korrigiert und ergänzt werden. Die nach dieser redaktionellen Überarbeitung erstellten Lehrpfade werden langfristig, ebenfalls in dem beschriebenen Format, gespeichert und können den Studierenden zum Lernen zugänglich gemacht werden.

4.5 Analyse der Inhalte des WiBA-Net

Neben den eigentlichen Inhalten des WiBA-Net interessieren im Rahmen dieser Arbeit vor allem die statistischen Daten der Texte dieser Inhalte. Als Grundlage für textstatistische und computerlinguistische Berechnungen werden Informationen über die Häufigkeiten von Begriffen, ihre Zerlegung in Komposita und ihre Wortstämme benötigt, die durch das beschriebene Verfahren zur Textaufbereitung gewonnen werden können. Weitere Daten beziehen sich auf den Textkorpus des WiBA-Net selbst.

Die Neukombination von Wissenseinheiten zu Lerneinheiten wird im Rahmen der vorliegenden Arbeit ausschließlich auf der Basis der in diesen Wissenseinheiten enthaltenen Texte und textstatistischen Informationen durchgeführt. Aus diesem Grund werden die für die Wissenseinheiten im Lernnetzwerk WiBA-Net vergebenen Metadaten nicht verwendet.

Die Inhaltsbasis des Lernnetzwerks WiBA-Net umfasst 2.394 Wissenseinheiten⁴¹. Werden aus diesen alle doppelten, leeren, nur aus Multimedia-Objekten bestehenden oder fehlerhaften Dateien ausgesondert, so erhält man 2.040 separate, inhaltstragende Wissenseinheiten. Diese Wissenseinheiten enthalten 219.980 Worte. Hieraus lässt sich abschätzen, dass eine Wissenseinheit im Schnitt ca. 108 Worte enthält. Die Gesamtanzahl der Worte setzt sich aus 28.053 verschiedenen Termen zusammen. Die Anzahl der verschiedenen Terme bezieht sich auf die Anzahl vor dem Stemming, also der Rückführung der einzelnen Terme auf ihre grammatische Grundform. Durch Stemming lässt sich die Anzahl der unterschiedlichen Terme noch einmal auf 19.347 reduzieren. Diese Wortmenge bildet das so genannte „Vokabular“ des WiBA-Net Textkorpus. Es finden sich 10.567 Wortformen, die durch Anfügen von Endungen von einer von 4.807 Stammformen abgeleitet wurden. Die restlichen Begriffe treten nur in einer grammatischen Form auf, können aber durchaus mehrfach vorhanden sein.

41 Stand: 19.04.2006

Die Verteilung der Terme kann Abbildung 4.4 entnommen werden. Das Diagramm stellt die Anzahl der Terme in Abhängigkeit der Häufigkeit ihres Vorkommens dar. Man liest z.B. ab, dass es 9.446 Terme gibt, die genau einmal vorkommen. Generell gilt, dass für geringere Häufigkeiten mehr Begriffe diese Häufigkeit aufweisen.

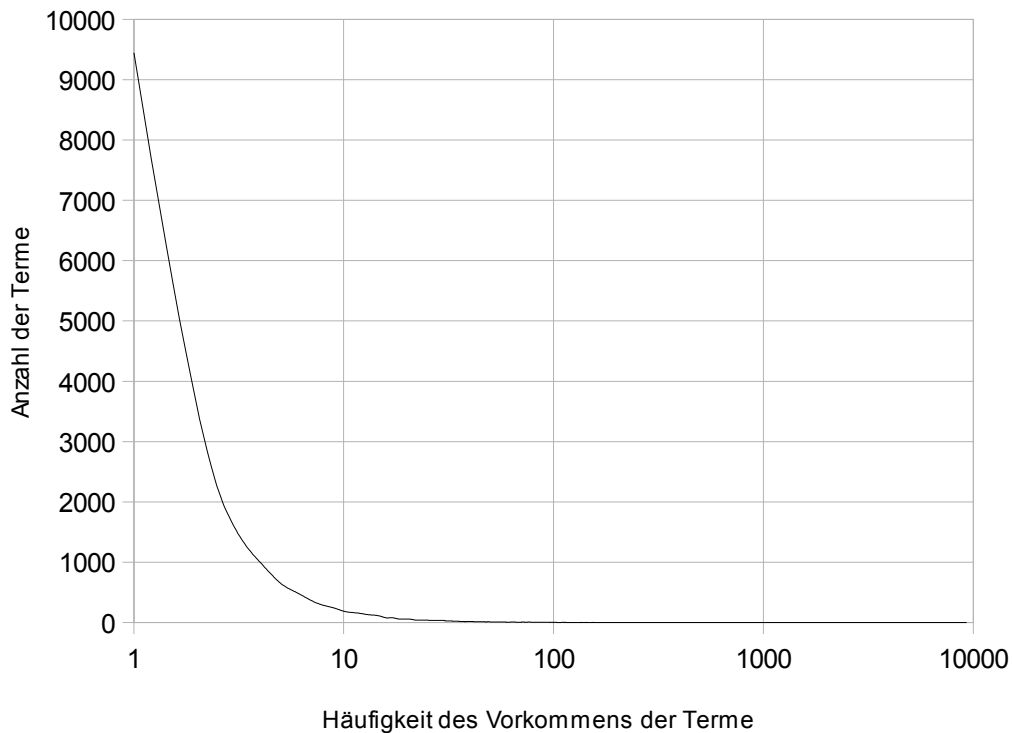


Abbildung 4.4: Worthäufigkeiten im WiBA-Net (logarithmischer Maßstab)

Sehr häufig vorkommende Begriffe sind als Suchterme nur bedingt geeignet, da mit ihnen eine sehr große Ergebnismenge produziert wird. Die am häufigsten vorkommenden Worte werden als Stoppworte deshalb von der Suche ausgeschlossen. Beispielhaft sind in Anhang D die im WiBA-Net am häufigsten vorkommenden Begriffe dargestellt. Zu selten vorkommende Terme sind ebenfalls schlecht als Suchbegriffe geeignet, da man diese Terme als Suchender kennen muss um eine sinnvolle Suchanfrage formulieren zu können. Die Terme kommen zudem mit hoher Wahrscheinlichkeit nicht in allen für die Suchanfrage relevanten Dokumenten vor. Wie in der Praxis üblich, werden die seltenen Terme aber nicht aus dem Suchkatalog entfernt.

Nach dem Entfernen der häufigsten Begriffe („Stoppworte“) und dem Ausschließen aller sich daraus ergebenden Sätze, die nur noch aus einem Wort bestehen würden, ergibt sich eine Anzahl der Sätze für die weitere Berechnung von 14.682. Bei der Kompositazerlegung lassen sich 9.424 Terme identifizieren, die sich aus anderen Begriffen des WiBA-Net Korpus zusammensetzen.

Eine wichtige Regel aus der Computerlinguistik ist das Zipfsche Gesetz (vgl. z.B. [FERBER 2003], [HEYER ET AL. 2006]). Es besagt, dass die Häufigkeit eines Terms in einem Textkorpus umgekehrt proportional zu dem sich aus dieser Häufigkeit ergebenden Rang des Terms ist. Für den WiBA-Net Korpus ergibt sich die in Abbildung 4.5 dargestellte Kurve, wenn man die Häufigkeit der Terme in einem doppelt-logarithmischen Maßstab über ihrem Rang aufträgt. Die Werte folgen im Wesentlichen einem linearen Verlauf, wie durch die eingezeichnete Regressionsgerade veranschaulicht wird. Abweichungen von diesem Zusammenhang ergeben sich, wie auch in der Abbildung zu erkennen ist, nur für sehr häufige oder sehr seltene Wörter.

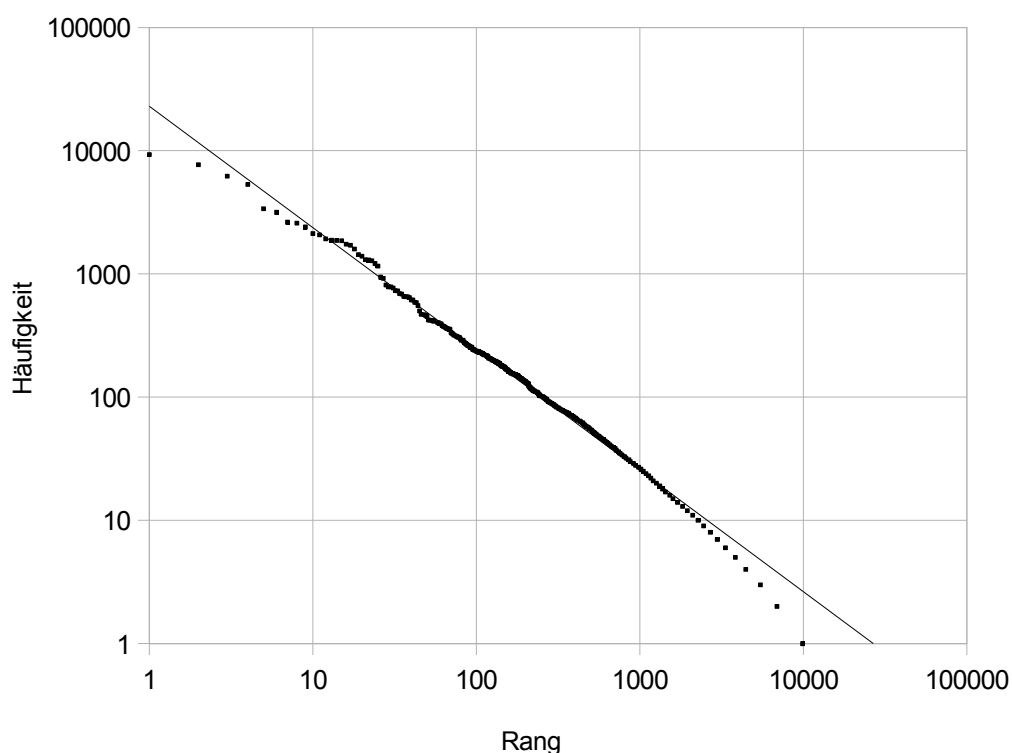


Abbildung 4.5: Darstellung der Häufigkeit der vorhandenen Terme in Abhängigkeit ihres Ranges (Zipfsches Gesetz)

Das Zipfsche Gesetz kann genutzt werden, um verschiedene Eigenschaften eines Textkorpus abzuschätzen, u.a. die Größe von dessen Vokabular. Nach HEYER ET AL. 2006 besteht ein Zusammenhang zwischen der Gesamtwortanzahl eines Textkorpus und dem Umfang des Vokabulars dieses Textkorpus, der sich durch eine Potenzfunktion ausdrücken lässt. Kleineren Textkorpora müssen weniger neue Dokumente zugefügt werden, damit sich in diesen mit hoher Wahrscheinlichkeit neue, noch nicht im Vokabular des Textkorpus befindliche Begriffe befinden. Ist ein Textkorpus bereits sehr umfangreich, so werden noch nicht im Vokabular befindliche Begriffe beim Hinzufügen neuer Dokumente nur selten gefunden. Der Zusammenhang zwischen dem Umfang N eines Textkorpus und der Größe t seines Vokabulars lässt sich durch Formel (4.3) ausdrücken.

$$t = a \cdot N^b \quad (4.3)$$

Die Parameter a und b dieser Gleichung sind abhängig vom betrachteten Textkorpus. Sie können durch eine Regressionsrechnung bestimmt werden. Dazu werden die im WiBA-Net enthaltenen Dokumente in der Reihenfolge ihrer Erstellung⁴² daraufhin durchsucht, ob in ihnen neue Begriffe vorkommen, die in den vorigen Wissensseinheiten noch nicht enthalten waren. Das Ergebnis dieser Auswertung ist in Abbildung 4.6 dargestellt. Man erkennt, dass sich der Vokabularzuwachs mit wachsendem Umfang des Textkorpus abschwächt.

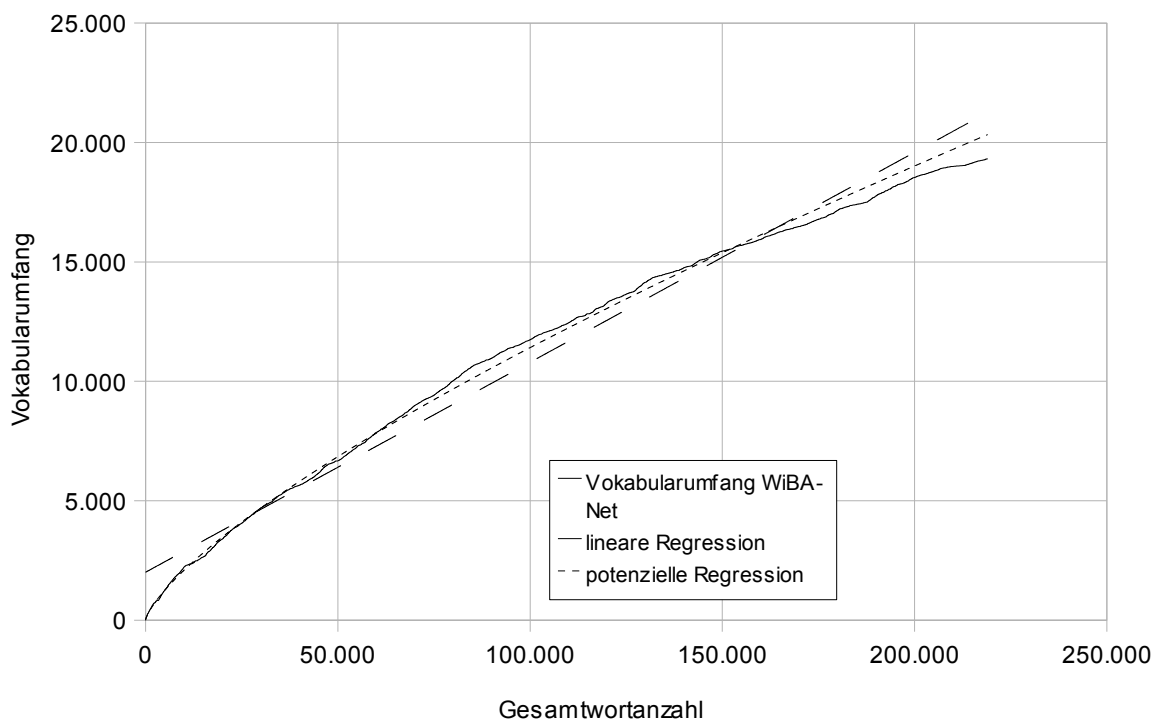


Abbildung 4.6: Abhängigkeit des Vokabularumfangs des WiBA-Net von der vorhandenen Gesamtwortanzahl

Nach Durchführen der Regressionsrechnung [PAPULA 1994] ergibt sich für den Umfang des Vokabulars des WiBA-Net folgende Gleichung:

$$t = 2,36 \cdot N^{0,74} \quad (4.4)$$

Da es sich beim WiBA-Net-Korpus um einen vergleichsweise kleinen Textkorpus handelt, weist der rechte Teil des Graphen immer noch eine merkbare Steigung auf. Des Weiteren weist die zum Vergleich in Abbildung 4.6 eingetragene lineare Regressionsgerade keine großen Unterschiede zu den Werten der Potenzfunktion auf. Das bedeutet, dass bei neu hinzukommenden Wissensseinheiten

⁴² Die Reihenfolge kann aus dem fortlaufenden Dateinamen abgeleitet werden.

ten noch relativ viele unbekannte Begriffe gefunden werden, ein Umstand, der vor dem Gesichtspunkt der Erweiterbarkeit des Systems berücksichtigt werden muss.

4.6 Erweiterung des Systems

Das Lernnetzwerk WiBA-Net war bei Projektende im Sinne des Hochschullehrermemorandums inhaltlich vollständig. Jedoch können Aktualisierungen und Ergänzungen des Systems prinzipiell⁴³ jederzeit vorgenommen werden. Es stellt sich deswegen die Frage, wie solche Änderungen des Textkorpus möglichst effizient in den nach der oben beschriebenen Vorgehensweise gewonnenen textstatistischen Informationen berücksichtigt werden können. Ferner ist zu definieren, wie weitere, bestehende E-Learning-Inhalte aus anderen Wissensgebieten möglichst einfach zu einer der vorgestellten Form analogen Inhaltsbasis aufbereitet werden können. Wird eine neue Wissensdomäne erst noch in E-Learning-Inhalte umgesetzt, so ist außerdem zu berücksichtigen, wie dies im Hinblick auf eine möglichst effiziente Einbeziehung in eine entsprechende Inhaltsbasis geplant werden kann.

Das grundlegende Problem bei der Erweiterung der Inhalte, die in der beschriebenen Weise ausgewertet werden, ist, dass durch fast alle Änderungen oder Erweiterungen zunächst einmal die Basis der statistischen Berechnungen geändert wird. Diese ist abhängig von der Anzahl der im System enthaltenen Begriffe und Sätze und der Anzahl der Kookkurenzen der Begriffe. Man muss also die textstatistischen Informationen nicht nur für neu hinzugekommene Terme berechnen, sondern wegen des Einflusses der Anzahl der Sätze im Textkorpus grundsätzlich alle Berechnungen auf der Basis einer neuen Grundgesamtheit neu durchführen. Diese Berechnungen in direktem Zusammenhang mit der Dateneingabe in Echtzeit durchzuführen, ist nicht praktikabel. Dies liegt an den teilweise sehr langen Rechenzeiten, die bei der Bestimmung der Signifikanzwerte der Kookkurenzen oder bei der Bestimmung ähnlicher Begriffe auftreten. Da die Anzahl der Sätze mit 14.682 bezogen auf die pro Zeiteinheit hinzukommenden Sätze jedoch relativ groß ist, kann auf eine sofortige Neuberechnung zugunsten einer in regelmäßigen Zeitabständen durchgeführten Neuberechnung verzichtet werden.

Ebenfalls berücksichtigt werden muss auch eine etwaige Löschung von Dateien oder eine Aktualisierung, bei der einzelne Begriffe aus Wissensseinheiten entfernt werden. Der Einfluss der darin enthaltenen Begriffe, die jetzt nicht mehr Bestandteile des Systems sind, muss aus den Ergebnissen aller textstatistischen Berechnungen entfernt werden. Bei einer regelmäßigen Neuberechnung aller Daten gehen diese Änderungen aber automatisch in das Berechnungsergebnis ein.

43 im Rahmen von gemeinsam beschlossenen Richtlinien

Durch neu in das System eingebrachte oder geänderte Wissensseinheiten werden Verknüpfungen von Termen geändert. Diese können erst dann berücksichtigt werden, wenn die relevanten textstatistischen Informationen – z.B. nach festen Zeitabschnitten – aktualisiert worden sind. Gefunden werden können neue oder geänderte Seiten schon früher, da die Vektordateien⁴⁴ und die inverse Dokumentdatei bei einem Neueintrag oder einer Änderung in Echtzeit geändert werden können. Die Frage einer inkrementellen Neuberechnung nur für neu hinzugekommene Daten erübrigt sich durch die sowieso erforderliche regelmäßige vollständige Neuberechnung.

Durch den im vorigen Abschnitt beschriebenen Zusammenhang zwischen der Mächtigkeit eines Textkorpus und dem Umfang seines Vokabulars kann abgeschätzt werden, mit welcher Frequenz unbekannte Worte bei der Aufnahme neuer Wissensseinheiten in das System zu erwarten sind. Für eine Wortanzahl von $N \approx 220.000$ lässt sich mit Formel (4.4) abschätzen, dass für jeweils 14 Worte, die dem System hinzugefügt werden, ein bisher unbekanntes dabei ist. Legt man den o.a. durchschnittlichen Umfang einer Wissensseinheit von ungefähr 108 Worten zugrunde, so kann man grob schätzen, dass pro neuer Wissensseinheit ca. 8 neue Begriffe ins System aufgenommen werden. Die Inhaltsbasis unterliegt bei stetiger Pflege also auch hinsichtlich der Anzahl der Begriffe und der Kookkurenzen der Begriffe einer nicht zu vernachlässigenden Veränderungsrate. Folglich sollte die Aktualisierung bzw. Neuberechnung der textstatistischen Informationen der Inhaltsbasis möglichst häufig erfolgen.

Bei vorhandenen Inhalten eines E-Learning-Systems⁴⁵, die zur automatischen Lehrpfadgenerierung genutzt werden sollen, müssen keine grundsätzlichen Änderungen der Inhalte vorgenommen werden. Voraussetzung für die Nutzung der textlichen Inhalte ist lediglich, dass die Texte in einem Format vorliegen, das vom Rechner automatisch ausgelesen werden kann, sodass die Inhalte von den formatierenden Elementen getrennt werden. Ausgangspunkt für die allen Inhaltsbasen gemeinsame Weiterverarbeitung ist die in Abschnitt 4.2.1 beschriebene Listendatei, die die Inhalte aller zugehörigen Wissensseinheiten absatzweise auf Zeilen enthält. Insofern ist nur der erste Teil des Konvertierungsvorgangs („Auslesen der Inhalte“) anzupassen, wenn sich die Inhalte von dem beschriebenen HTML-Format unterscheiden. Liegen die Inhalte vom Format getrennt vor, z.B. in einer Datenbank oder in einem XML-Format, so vereinfacht sich dieser Schritt entsprechend. Nach der Erstellung dieser Listendatei können die weiteren Schritte wie beschrieben für die neue Domäne durchgeführt werden, sodass diese zusätzlich zu ihrer ursprünglichen Nutzung zur automatischen Lehrpfadgenerierung benutzt werden kann.

44 Nr. XI nach Tabelle 4.1

45 Es wird davon ausgegangen, dass die zu nutzenden Inhalte einen ähnlichen Aufbau haben wie das WiBA-Net. Insbesondere darf jede Wissensseinheit nur einen Inhalt umfassen und muss kontextfrei sein.

Wissenseinheiten, die in einem dem WiBA-Net entsprechenden Format vorliegen, können ohne Änderung des Algorithmus konvertiert werden. Liegt der Textkorpus in einem anderen Format vor, muss auf jeden Fall sichergestellt sein, dass der Algorithmus erkennen kann, wo neue Wissenseinheiten beginnen. Auf dieser Basis lässt sich das Auslesemodul des Algorithmus dann anpassen.

Die Wissenseinheiten als solche müssen vorhanden sein. Sie brauchen nicht in eigenen Dateien zu stehen, jedoch muss der jeweilige Text formuliert und mit einer als solche identifizierbaren Überschrift versehen sein. HTML-Dateien für die Anzeige der einzelnen Wissenseinheiten können beim Auslesen automatisch gespeichert werden.

Neue E-Learning-Inhalte sollten nicht für eine ausschließliche Nutzung im Rahmen einer automatischen Lehrpfadgenerierung geplant werden. Diese kann erst bei umfangreicheren Inhaltsbasen sinnvoll funktionieren, da für aussagekräftige textstatistische Informationen größere Datenmengen benötigt werden. Mit den textstatistischen Informationen können dann auf die beschriebene Weise Beziehungen zwischen Dokumenten oder Begriffen hergestellt werden. Anfänglich sind neue Inhaltsbasen für das Finden von Zusammenhängen zwischen Begriffen oder Dokumenten noch zu klein. Es empfiehlt sich also, zunächst manuell zu erstellende Lehrpfade zu planen, und dem Nutzer ab einem bestimmten Umfang der Inhaltsbasis die zusätzliche Nutzung einer automatischen Lehrpfadgenerierung zu ermöglichen. Jedoch kann die Datenverwaltung einer neu als E-Learning umzusetzenden Wissensdomäne von Beginn an optimiert werden, indem die Inhalte strukturiert niedergelegt werden. Die Nutzung einer Datenbank oder eines XML-Formats zur Speicherung der Inhalte bietet sich an.

Theoretisch können auch bestehende Lehrpfade in eine Erweiterung mit einbezogen werden. Wenn der Hochschullehrer z.B. regelmäßig die Inhaltsbasis durch aktuelle Inhalte ergänzt, können diese Inhalte, anstatt nur mit einzelnen Wissenseinheiten zu neuen Lerneinheiten verknüpft zu werden, auch da, wo es sinnvoll ist, an bestehende, manuell erstellte WiBA-Net Lehrpfade angefügt werden. Hierdurch kann auch „Just-In-Time“-Lernen sinnvoll unterstützt werden. Grundlegende Informationen können so auf einem aktuellen Stand gehalten werden, was sich insbesondere in der Weiterbildung anbietet.

Zunächst muss die Information, welche Lehrpfade welche Wissenseinheiten enthalten, ausgewertet werden. Auf dieser Basis wird dann die Ähnlichkeit einer „aktuellen“ Seite zum Gesamtthema eines Lehrpfads bestimmt. Wenn diese Ähnlichkeit einen bestimmten Wert überschreitet, kann diese Seite dem Lehrpfad hinzugefügt werden. Sie muss natürlich als automatisch hinzugefügte Seite gekennzeichnet werden. Die hinzugefügte Seite ist als Nebenseite in das passendste Cluster einzuordnen. Die Option, bestehende Lehrpfade zu ändern, sollte in jedem Fall fakultativ sein, da auf diese Weise manuell erstellte Inhalte verändert werden und nicht auszuschließen ist, dass eine Fehlzusammenordnung auftritt. Deshalb muss der Nutzer bestimmen können, ob solchermaßen modifizier-

te Lehrpfade angezeigt werden, entweder der Lerner selbst, oder der Lehrende für seine Lerner. Möglich wäre natürlich auch eine manuelle Überprüfung der automatischen Zuordnung und eine anschließende Freigabe durch den Hochschullehrer.

5 Automatische Generierung von Lehrpfaden

5.1 Grundlagen der Lehrpfaderstellung

5.1.1 Bedeutung für das didaktische Konzept des WiBA-Net

Der Vorschlag zur automatischen Generierung von Lehrpfaden greift auf die im Projekt WiBA-Net entwickelten Konzepte und gewonnenen Erfahrungen zurück. Dazu zählt insbesondere eine der Kernfunktionalitäten des WiBA-Net: die Lehrpfade als Darstellung des zu vermittelnden Wissens in strukturierten Einheiten. Diese werden im WiBA-Net hauptsächlich dazu genutzt, Studierenden im Grundstudium Fachwissen zu vermitteln. Dabei werden von den Autoren manuell erstellte Lehrpfade eingesetzt. Im Rahmen des vorgeschlagenen Systems sind zusätzlich individuell für einen Anwender generierte Lehrpfade nutzbar. Diese Lehrpfade erweitern das didaktische Konzept des WiBA-Net durch zusätzliche Einsatzmöglichkeiten für dessen Inhalte.

In Tabelle 5.1 findet sich ein Überblick der im WiBA-Net vorhandenen Funktionalitäten zur Unterstützung von Vorlesungen. Diese wurden in Abschnitt 3.6 vorgestellt und werden an dieser Stelle um die Funktionalitäten ergänzt, die sich durch die automatische Lehrpfadgenerierung ergeben. Die neuen Funktionalitäten werden zusätzlich in das didaktische Konzept des WiBA-Net übernommen. Sie können sowohl von Hochschullehrern als auch von Studierenden zur Vor- und Nachbereitung von Vorlesungen genutzt werden.

Nach dem bestehenden Konzept des WiBA-Net bearbeiten Studierende in der Vorbereitungsphase im Selbststudium vom Hochschullehrer empfohlene Lehrpfade, sodass in der nächsten Vorlesung eine Vertiefung des Themas und eine Diskussion stattfinden kann. Durch die automatisch erstellten Lehrpfade kann vom Lehrenden auch mit konkreten Fragestellungen gearbeitet werden, ohne spezielle Lehrpfade angeben zu müssen. So kann den Lernenden ein aufgabenorientiertes Lernen ermöglicht werden. Die Fragestellungen können zunächst explizit oder auch durch eine Aufgabenstellung implizit vorgegeben sein. Mit steigendem Wissen des Lernalters kann eine durch diese Aufgaben erzeugte extrinsische Motivation durch eine intensive Auseinandersetzung mit dem Thema und die Entwicklung von Interesse daran in eine intrinsische Motivation übergehen [WÖRNER 2006].

In der Nachbereitungsphase können von Studierenden Verständnisprobleme, die nach der Präsenzveranstaltung offen geblieben sind, mit Hilfe des Systems gelöst werden. Begriffe und Konzepte, die nicht verstanden wurden, können als Suchanfrage benutzt werden. Aus dem generierten Lehrpfad können Hintergründe und Zusammenhänge erschlossen werden, da auch der Kontext der unbekannten Begriffe bei der Zusammenstellung berücksichtigt wird. So können Lernende ge-

zielt Teilgebiete vertiefen, für die sie sich persönlich besonders interessieren. Das Themenfeld einer Suchanfrage kann auf der Basis eines Lehrpfads selbstständig erarbeitet werden. Bei Bedarf können durch weitere Suchanfragen zusätzliche Lehrpfade zu verwandten Themen erstellt werden. Auf diese Weise lassen sich gezielt Lücken im Faktenwissen des Lernalers schließen. Nach demselben Prinzip lassen sich die automatisch generierten Lehrpfade auch noch in späteren Semestern als Nachschlagewerk nutzen.

*Tabelle 5.1: Ablaufschema von Vorbereitungsphase, Präsenzveranstaltung und Nachbereitungsphase
(Funktionalitäten des bestehenden WiBA-Net in Normalschrift, Anwendungen der automatischen Lehrpfadgenerierung in Kursivschrift)*

Nutzer	Vorbereitungsphase	Präsenzveranstaltung	Nachbereitungsphase
Hochschullehrer	<ul style="list-style-type: none"> x Planung der Vorlesung x Zusammenstellung von Materialien x <i>Spezielle Themen:</i> <ul style="list-style-type: none"> - <i>automatisch generierte Lehrpfade als Grundlage</i> 		<ul style="list-style-type: none"> x Nachbearbeitung x Einfügen von Links und Materialien x Benennung der Inhalte für die nächste Vorlesung x Freischaltung
Hochschullehrer / Studierende		Vorlesung inkl. <ul style="list-style-type: none"> x Vortrag x Tafelanschrift x Präsentation von Materialien 	
Studierende	<ul style="list-style-type: none"> x Abrufen der Inhaltsangabe der nächsten Vorlesung x Selbststudium: <ul style="list-style-type: none"> - empfohlene Lehrpfade (Grundlagen) - <i>automatisch generierte Lehrpfade</i> 		<ul style="list-style-type: none"> x Download Präsentation inkl. Tafelanschrift x <i>Verständnisprobleme:</i> <ul style="list-style-type: none"> - <i>automatisch generierte Lehrpfade</i> x Wiederholen und Vertiefen, Übungen, Links, Tests

Hochschullehrer können in der Vor- oder Nachbereitungsphase einer Vorlesung mit Hilfe des vorgeschlagenen Systems Lehrpfade zu bestimmten Themen erstellen, die dann von den Studierenden als Vorbereitung für die nächste Veranstaltung bearbeitet werden. Es wird dadurch möglich, schnell vertiefende Lehrpfade zu Themen, die in der Diskussion stehen, zu erstellen. Im Unterschied zu den Lernenden hat ein Hochschullehrer bei der Benutzung des Systems die Möglichkeit, die automatisch erstellten Lehrpfade manuell nachzubearbeiten und nach eigenem Ermessen zu ändern, zu ergänzen oder zu kürzen. Die didaktischen und inhaltlichen Vorstellungen des Lehrenden werden durch diese Bearbeitungsmöglichkeit weitgehend unterstützt. Der entfallende Aufwand

für Zusammenstellung und Gliederung der grundlegenden Inhalte kann somit eine Arbeitserleichterung für den Lehrenden darstellen.

5.1.2 Aufbau des Systems

Das vorgeschlagene System erlaubt die automatische Erstellung von Lehrpfaden auf der Basis einer Nutzeranfrage in Form von Suchbegriffen. Zur Erstellung werden ausschließlich aus den Inhalten gewonnene Zusammenhänge verwendet. Damit ein Lehrpfad erstellt werden kann, muss der Nutzer seine Anfrage in Form von Suchbegriffen formulieren und in das System eingeben. Aus diesen Suchbegriffen wird dann mit aus dem Text gewonnenen Informationen ein die Thematik der einzelnen Suchbegriffe übergreifendes Themenfeld aus „verwandten“ Begriffen erstellt. Die Begriffe dieses Themenfeldes dienen zur Erweiterung der Suchterme. Damit ein solches Themenfeld sinnvoll zusammengestellt werden kann, ist es erforderlich, dass immer mindestens zwei Suchbegriffe verwendet werden.

Mit Hilfe der erweiterten Suchterme werden die passenden Wissensseinheiten im System ausgewählt. Diese werden mittels textstatistischer und linguistischer Informationen nach ihrer Relevanz beurteilt. Dabei werden zusätzliche Themenschnittpunkte zwischen jeweils zwei Stichworten gesondert berücksichtigt. Die relevantesten Seiten werden abschließend sortiert und gegliedert, so dass eine didaktisch möglichst sinnvolle Anordnung entsteht. Der fertig erstellte Lehrpfad wird dem Nutzer angezeigt.

Das entwickelte System zur automatischen Lehrpfadgenerierung besteht aus vier aufeinander aufbauenden Stufen:

1. Erstellung des Suchvektors,
2. Suche nach Wissensseinheiten mit Hilfe des Suchvektors,
3. Bestimmung der Wissensseinheiten, die am besten zum Thema der Suche passen,
4. Sortierung und Gliederung der erhaltenen Wissensseinheiten und Erstellen des Lehrpfads.

Es werden während des Berechnungsablaufs unterschiedliche Zwischenergebnisse erzielt, die zusammen mit der Datenbank der textstatistischen und linguistischen Informationen jeweils Grundlage für die nächste Stufe sind. Der Ablauf ist in Abbildung 5.1 schematisch dargestellt und wird in Abschnitt 5.1.3 erläutert.

Die Zusammenstellung von Wissensseinheiten zu didaktisch sinnvollen Lehrpfaden beruht im Wesentlichen auf der Nutzung von textstatistischen und linguistischen Informationen, die mit geeigneten Verfahren direkt aus den Texten der Wissensseinheiten gewonnen werden. Die diesbezügliche Aufbereitung und Auswertung der Texte wurde bereits in Kapitel 4 beschrieben. Aus den gewonnenen Informationen können Rückschlüsse über die semantischen Verknüpfungen zwischen Texten

oder zwischen Begriffen gezogen werden. An dieser Stelle werden die zur Verfügung stehenden Informationen zusammengefasst und dabei ihrer in diesem Kapitel beschriebenen Anwendung zugeordnet. Eine Übersicht findet sich in Tabelle 5.2.

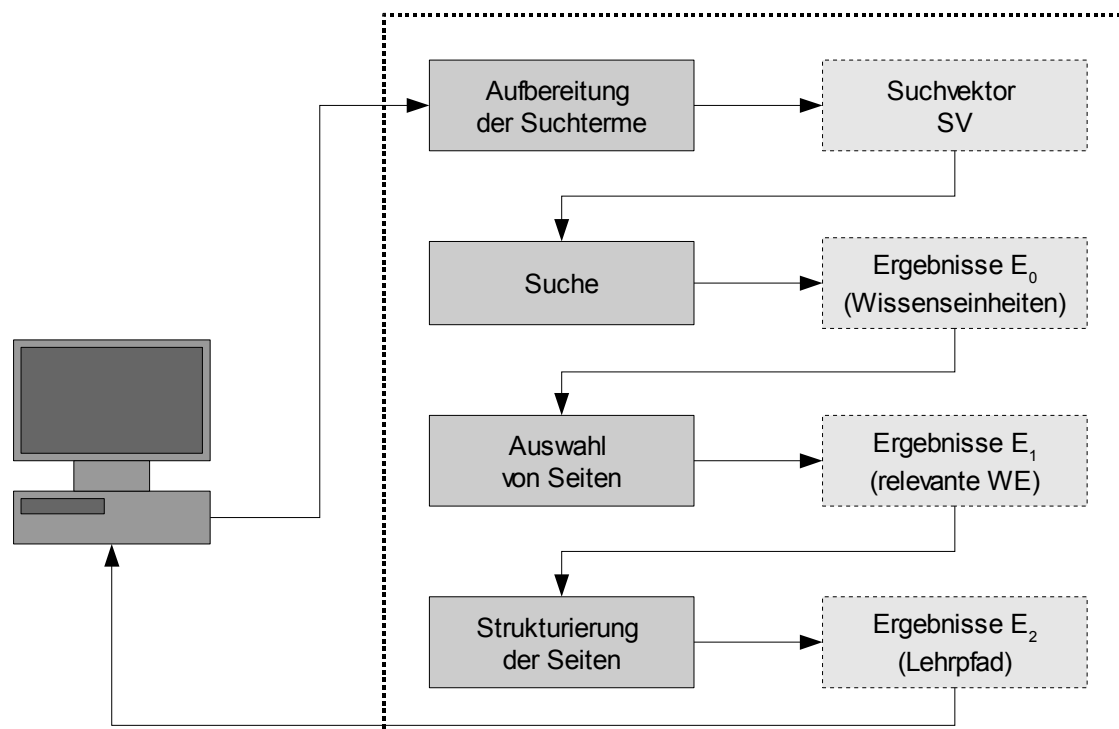


Abbildung 5.1: Schematische Darstellung des vierstufigen Ablaufs der Lehrpfadgenerierung

Tabelle 5.2: Vorhandene textstatistische und linguistische Informationen und ihre Nutzung

Textstatistische und linguistische Informationen	Grundlage für welches Verfahren
Zuordnung Ableitung → Stammform	Suchtermerweiterung
inverse Liste (Wort → Datei)	Boolesche Suche
Liste aller Worte in Stammform	Vergleich mit anderen Termen
Häufigkeiten der Worte	statistische Berechnungen Berechnung von Ähnlichkeitsmaßen Sortierung
Zuordnung Komposita → Konstituenten	Suchtermerweiterung Berechnung von Ähnlichkeitsmaßen

Textstatistische und linguistische Informationen	Grundlage für welches Verfahren
	Sortierung
Vereinigung der Kookkurenzlisten	Suchterweiterung Berechnung von Ähnlichkeitsmaßen Relevanzbestimmung Bestimmung von Fachausdrücken Sortierung
Ähnlichkeitswerte für Begriffspaare	Suchterweiterung Relevanzbestimmung Bestimmung weiterführender Seiten
Vektordateien	Berechnung von Ähnlichkeitsmaßen Sortierung Bestimmung von Häufigkeiten in einzelnen Dateien

5.1.3 Ablauf des Algorithmus

5.1.3.1 Schritte zur Lehrpfaderstellung

Am Beispiel der Inhalte des WiBA-Net wird in den folgenden Abschnitten der Ablauf des prototypischen Programms detailliert erläutert. Dabei wird insbesondere auf die bei den einzelnen Schritten zu bestimmenden Zwischenergebnisse eingegangen. An dieser Stelle wird zunächst der Gesamtprozess im Überblick vorgestellt. Der tatsächliche Ablauf der Berechnungsschritte ist anhand exemplarischer Lehrpfade in Kapitel 6 dargestellt. Die Auswertung des ersten der dort vorgestellten Lehrpfade enthält eine umfassende Dokumentation aller Schritte.

Die mit dem vorgeschlagenen System erstellten Lehrpfade beruhen auf vom Nutzer eingegebenen Suchbegriffen. Es wird dabei unterstellt, dass der Nutzer pro Suchvorgang bzw. pro Lehrpfaderstellung Informationen benötigt, die ein die Suchbegriffe umfassendes und übergreifendes Thema betreffen. Deshalb sind immer mindestens zwei Suchbegriffe erforderlich. Alle eingegebenen Suchbegriffe werden vom System gleichrangig behandelt. Aus ihnen wird ein Lehrpfad erstellt, dessen Wissensseinheiten möglichst zu allen durch die Suchworte vorgegebenen Themen Bezüge haben⁴⁶.

Nachdem der Nutzer das abzudeckende Themenfeld mit Hilfe von Suchbegriffen definiert hat, werden die vorgegebenen Suchbegriffe mit Begriffen, die aus den textstatistischen Informationen erschlossen werden, erweitert. Mit dem erweiterten Suchvektor SV wird eine Boolesche Suche über

⁴⁶ Die übergreifende Thematik wird später noch durch Unterthemen ergänzt, die aus Kombinationen einzelner Suchbegriffe abgeleitet werden (vgl. Abschnitt 5.3.5).

alle Wissensseinheiten durchgeführt. Diese Suche mit erweiterten Suchtermen dient der Erzeugung einer Ausgangsmenge von Wissensseinheiten E_0 mit möglichst hohem Recall-Wert, aus der im Lauf der weiteren Verarbeitung die für das Thema bzw. den Lehrpfad relevanten Seiten identifiziert werden. Demgegenüber ist der Precision-Wert der Ausgangsmenge E_0 relativ niedrig, da im ersten Schritt auch viele Seiten gefunden werden, die mit dem eigentlichen Thema nichts zu tun haben.

Für die gefundenen Wissensseinheiten der Menge E_0 werden anschließend Ähnlichkeitsberechnungen mit dem Suchvektor und weiteren, auf den Suchbegriffen basierenden Vektoren durchgeführt. Die berechneten Ähnlichkeiten werden zusammen mit Informationen über das anteilige Vorkommen der ursprünglichen Suchbegriffe im Text und in der Überschrift der Wissensseinheiten dazu verwendet, mittels Fuzzy Logic für jede Seite einen Relevanzwert bezüglich der Zugehörigkeit zum Lehrpfad zu bestimmen. Die Wissensseinheiten mit den höchsten Relevanzwerten werden ausgewählt und abschließend um solche Wissensseinheiten bereinigt, die einen besonders geringen Bezug zu einzelnen Suchbegriffen oder Kombinationen einzelner Suchbegriffe haben. Sie werden um solche Wissensseinheiten ergänzt, die einen entsprechenden Bezug haben, der besonders hoch ist. Die verbleibenden Wissensseinheiten bilden die Ergebnismenge E_1 und werden im nächsten Schritt weiterverarbeitet.

Aus den in der Ergebnismenge E_1 verbleibenden Wissensseinheiten wird nun ein strukturierter Lehrpfad zusammengestellt. Zunächst werden einander ähnliche Wissensseinheiten zu Clustern zusammengefasst. Danach wird auf Basis didaktischer Gesichtspunkte die Reihenfolge der Cluster und ihre innere Strukturierung, also die Gliederung in Haupt- und Nebenseiten, festgelegt. Der fertige Lehrpfad stellt die Ergebnismenge E_2 dar.

Zwei Schritte, die nicht die Lehrpfaderstellung selbst betreffen, sondern vor- bzw. nachgeschaltet sind, werden im Folgenden erläutert. Dies sind zum einen die Längenbestimmung für den zu generierenden Lehrpfad und zum anderen die Anzeige des fertigen Lehrpfads.

5.1.3.2 Länge des Lehrpfads

Der Umfang eines zu erzeugenden Lehrpfads muss in erster Linie didaktischen Kriterien genügen. Zu kurze Informationen stellen den Lerner nicht zufrieden, und zu viele Informationen lenken von der zentralen Thematik ab und überfordern den Lerner. Natürlich müssen die im Lehrpfad enthaltenen Wissensseinheiten dabei eine hohe Relevanz für das Thema der Suchanfrage haben. Als Nebenbedingung soll die Struktur des Lehrpfads übersichtlich darstellbar sein. Deshalb wird für die Länge eines Lehrpfads der Sollwert L nach der im Folgenden beschriebenen Definition festgelegt.

Da nicht nur der ursprüngliche Suchbegriff, sondern auch Terme aus der Suchterweiterung zum Auffinden von Seiten führen, hängt die Anzahl der gefundenen Seiten E_0 von mehreren Variablen ab. Wenn im erweiterten Suchvektor sehr häufige Begriffe vorhanden sind, so steigt die Menge der gefundenen Seiten stark an. Sie kann also nicht als Grundlage für den Umfang des Lehrpfads herangezogen werden, da sie mit dem Thema der Anfrage in keiner direkten Beziehung steht⁴⁷. Es bietet sich deswegen an, den Sollwert L nur in Abhängigkeit von der Häufigkeit h der ursprünglichen Suchbegriffe zu berechnen.

Als Randbedingungen werden die minimale und die maximale Seitenanzahl eines Lehrpfads vorgegeben. Ein Lehrpfad sollte per Definition aus mindestens zwei Seiten bestehen, da ansonsten kein Pfad, sondern eine Einzelseite vorliegen würde. Die maximale Seitenanzahl wird einerseits aus der dem Lerner zumutbaren Bearbeitungsdauer bestimmt. Diese beträgt, analog zu den Lehrpfaden im WiBA-Net, ca. 30 bis maximal 45 Minuten. Die Bearbeitungsdauer der einzelnen Wissenseinheiten liegt zwischen einer und drei Minuten. Da ein zu wählender Maximalwert der Lehrpfadlänge nur von einigen Lehrpfaden erreicht wird, wird dieser Maximalwert im oberen Bereich der vorstehenden Randbedingungen mit 25 Seiten angenommen. Ein zusätzliches Kriterium für diese Wahl ist die übersichtliche Darstellbarkeit der Ergebnisse. Der Lerner soll zu jedem Zeitpunkt die Struktur des gesamten Lehrpfads und den aktuellen „Standort“ überblicken können. Für die gewählten 25 Seiten ist bei heute üblichen Bildschirmauflösungen⁴⁸ in der Regel die Darstellbarkeit der als Navigation untereinander angeordneten Seitentitel auf einer Bildschirmseite gewährleistet, ohne dass gescrollt werden muss⁴⁹.

Eine Gleichung für den Sollwert der Lehrpfadlänge kann aus der Grundüberlegung abgeleitet werden, dass das häufigste bedeutungstragende Stichwort den längsten möglichen Lehrpfad erzeugen soll. Das entspricht einem 25-seitigen Lehrpfad zu dem häufigsten Stichwort, nämlich „Stahl“ mit 693 Instanzen. Analog soll ein einmal vorkommendes Stichwort zu einem Lehrpfad von 2 Seiten Länge führen. Damit für mittlere Stichwörterhäufigkeiten keine zu kurzen Lehrpfade entstehen, wird eine logarithmische Funktion gewählt. Diese sorgt für einen raschen Anstieg der Lehrpfadlänge bei kleinen Häufigkeiten, im Bereich größerer Häufigkeiten wird die Steigung dafür geringer. Als Grundgleichung wird also $L = a \cdot \log(h) + b$ angesetzt, h bezeichnet dabei die Häufigkeit des entsprechenden Suchworts. Da der Logarithmus für $h = 1$ zu null wird, ergibt sich der Wert 2 für den Parameter b . Für $h = 693$ kann damit der Parameter a zu ungefähr 8 bestimmt werden. Für die Länge des Lehrpfads kann man also schreiben:

47 Aus demselben Grund eignet sich auch eine untere Grenze für den Relevanzwert nicht zum Festlegen einer Menge von Wissenseinheiten, da die den Grenzwert überschreitende Menge von Seiten ebenfalls abhängig von der Menge E_0 wäre.

48 Standardvorgabe im WiBA-Net waren 1024x768 Pixel.

49 Die tatsächlich darstellbare Anzahl hängt zusätzlich von der vom Lerner gewählten Schriftgröße und der Länge der Seitentitel ab, die ggf. einen Zeilenumbruch erfordern.

$$L = 8 \cdot \log(\bar{h}) + 2 \quad (5.1)$$

Der sich ergebende Sollwert für die Lehrpfadlänge L kann exemplarisch für verschiedene Begriffe Tabelle 5.3 entnommen werden.

Eine Suchanfrage wird immer als Wunsch nach Informationen zu dem Themenfeld verstanden, das die Begriffe der Suchanfrage umspannt. Es werden also nicht umfassende Informationen zu jedem Suchbegriff benötigt, sondern im Wesentlichen solche, die die Querschnittsthematik betreffen. Deshalb wird bei einer Suchanfrage aus mehreren Begriffen der Sollwert L auf Basis der Häufigkeit des seltensten Begriffs bestimmt. Die Wissenseinheiten werden nach dem in Abschnitt 5.3 beschriebenen Verfahren hinsichtlich ihrer Relevanz zum Thema des Lehrpfads bewertet. Die L relevantesten Seiten werden dann für den Lehrpfad berücksichtigt. Die Länge des Lehrpfads kann sich danach noch ändern, denn nach der Relevanzberechnung werden noch aus Suchwortpaaren abgeleitete Unterthemen berücksichtigt.

Tabelle 5.3: Exemplarische Suchbegriffe mit ihrer Häufigkeit im Korpus des WiBA-Net und die sich nach Formel (5.1) ergebenden Sollwerte für die Lehrpfadlänge L

Begriff	Häufigkeit h	Sollwert L
Konsistenzmaß	5	8 Seiten
Endkriechzahl	10	10 Seiten
Edelstahl	50	16 Seiten
Druckfestigkeit	107	18 Seiten
Eisen	194	20 Seiten
Holz	452	23 Seiten
Stahl	693	25 Seiten

Natürlich ist die Länge eines Lehrpfades außerdem von der Anzahl der zur Verfügung stehenden Wissenseinheiten abhängig. Sind zu einer Suchanfrage weniger als L objektiv relevante Wissenseinheiten vorhanden, bestehen die letzten Seiten des entsprechenden Lehrpfads demnach aus wenig oder gar nicht relevanten Seiten⁵⁰.

⁵⁰ Als Maßnahme gegen solche zu langen Lehrpfade könnten z.B. die am wenigsten relevanten Seiten entfernt werden, wenn diese gleichzeitig einen deutlich geringeren Relevanzwert als die nächst relevantere Wissenseinheit haben und ferner nur eine geringe Ähnlichkeit mit dem verbleibenden Rest des Lehrpfads aufweisen. Hierauf wird im Rahmen der vorliegenden Arbeit verzichtet, da nur wenige Lehrpfade betroffen sind und auch relevante Seiten fälschlich entfernt werden könnten. Dies kann anhand der Relevanzwerte R der in Anhang J analysierten Lehrpfade verifiziert werden. Wissenseinheiten am Ende eines Lehrpfads, die keine Relevanz für den entsprechenden Lehrpfad haben und deren Relevanz R zugleich einen überdurchschnittlich großen Abstand ($> 0,1$) zur nächsthöheren Relevanz hat, sind dort unterstrichen dargestellt. Insgesamt sind 4 von 62 untersuchten Lehrpfaden betroffen.

5.1.3.3 Anzeige der Lehrpfade

Im WiBA-Net wird das Programm *Learning Solution* der Firma SAP verwendet. Die Anzeige der automatisch generierten Lehrpfade mit diesem Programm ist zwar prinzipiell möglich. Jedoch muss zur Bearbeitung von Lehrpfaden auf dem Rechner eines Dozenten auf diesem die Autorenumgebung des Programms installiert werden. Diese ist zum einen lizenzpflichtig, zum anderen kann mit ihr nicht direkt mit auf dem Server erzeugten Lehrpfaddateien gearbeitet werden, sondern es müssen erst Kopien herunter geladen und nach der Bearbeitung wieder hoch geladen werden. Aus diesen Gründen eignet sich *Learning Solution* nur schlecht für die Anzeige der mit dem entwickelten System automatisch generierten Lehrpfade. Es kommt deshalb ein separat entwickeltes, serverbasiertes Programmmodul zur Lehrpfadanzeige zum Einsatz (vgl. Abbildung 5.2). Dieses kann die automatisch generierten Lehrpfade, die in einem auf XML basierenden Format vorliegen, einlesen und anzeigen. Für Autoren sind außerdem Bearbeitungsmöglichkeiten die Struktur und den Inhalt der Seiten betreffend vorhanden.

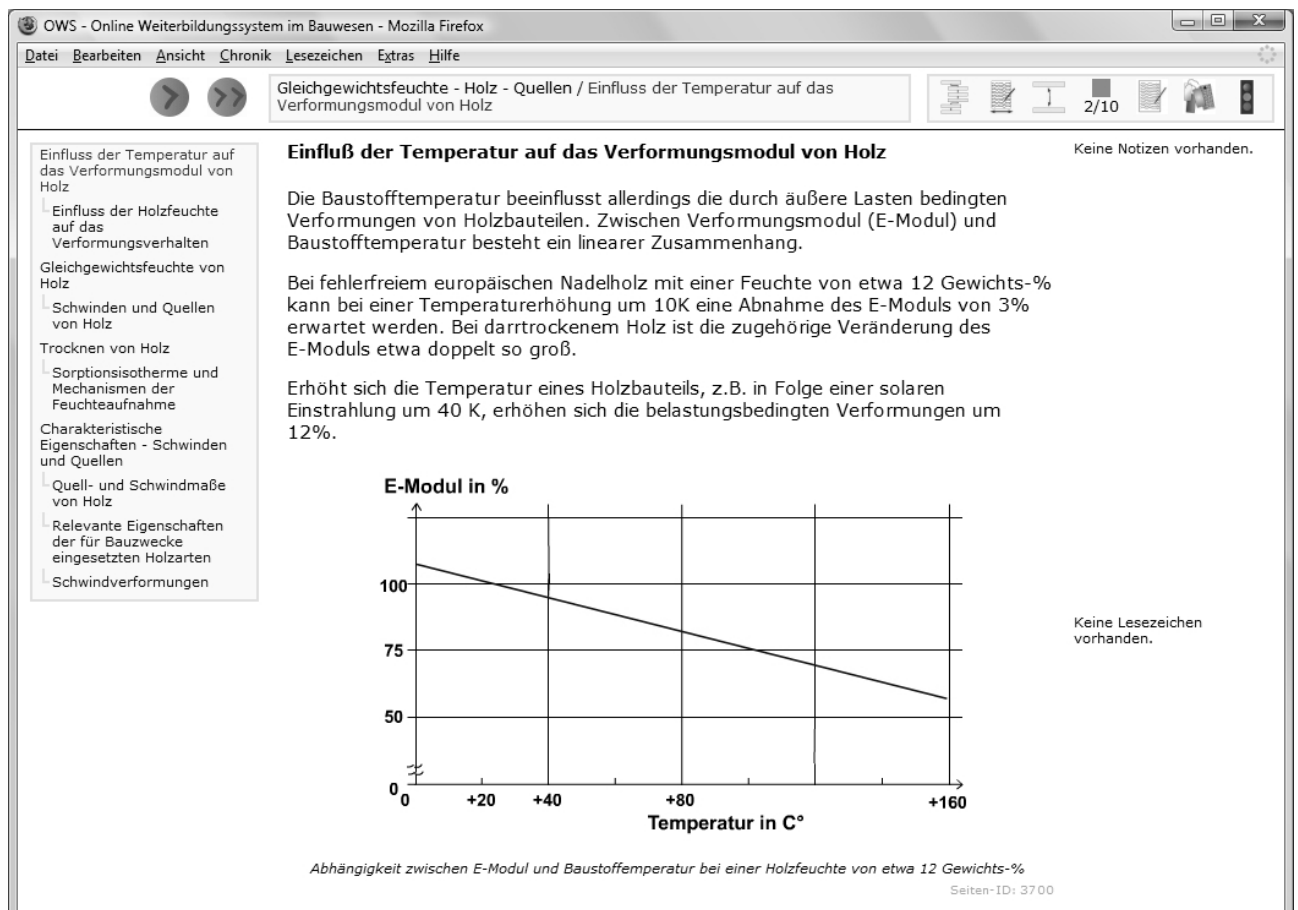


Abbildung 5.2: Ansicht des Anzeigeprogramms für die Lehrpfade

Das Anzeigeprogramm für die Lehrpfade gliedert sich in drei Bereiche. Links und oben findet sich der Navigationsbereich. Dieser erlaubt das Vor- und Zurückspringen sowie den direkten Zugriff auf beliebige Seiten. Eine Übersicht der Struktur des Lehrpfads befindet sich im grau hinterlegten Bereich auf der linken Seite. Der Inhaltsbereich zur Anzeige der aktuellen Wissensseinheit befindet sich in der Mitte des Bildschirms. Am rechten Rand findet sich ein Bereich für persönliche Informationen, in dem Notizen oder Lesezeichen zur aktuellen Wissensseinheit abgelegt werden können.

5.2 Suchanfrage

5.2.1 Suchtermerweiterung

Der Prozess der Lehrpfaderstellung beginnt mit der Eingabe der Suchbegriffe durch den Nutzer. Zunächst werden die eingegebenen Suchbegriffe normiert, d.h. Großbuchstaben und Umlaute werden ersetzt und die Begriffe werden, falls möglich, gestemmt. Dies ist notwendig, damit die Suchbegriffe mit den entsprechenden in den Inhalten und in den Textinformationen enthaltenen Wortformen der Inhaltsbasis⁵¹ übereinstimmen. Zu den vom Nutzer vorgegebenen Suchbegriffen werden alle im Laufe der weiteren Verarbeitung benötigten linguistischen und textstatistischen Informationen zusammengestellt. Diese werden bis zum Ende der Lehrpfaderstellung zwischengespeichert, damit in jedem Schritt auf sie zurückgegriffen werden kann.

Zu den benötigten linguistischen Informationen zählen die Wortstämme und Ableitungen der Suchbegriffe sowie mit ihnen gebildete Komposita bzw. ihre Konstituenten, falls es sich bei den Begriffen selbst um Komposita handelt. Dabei wird zusätzlich festgehalten, ob der jeweilige Begriff den ersten oder einen weiter hinten liegenden Platz in einem Kompositum einnimmt. Aus dieser Information können Rückschlüsse darüber angestellt werden, ob es sich um ein Determinativkompositum oder ein Kopulativkompositum handelt.

Als textstatistische Informationen werden zu allen Suchbegriffen die in der Kookkurenz-Datenbank gespeicherten Kookkurenzen ermittelt. Hier sind für alle bedeutungstragenden Terme des Textkorpus die aus Nachbar- und Satzkookkurenzen ermittelten vereinigten Kookkurenzen⁵² und deren Signifikanzwerte gespeichert.

Ferner lassen sich, wie in Abschnitt 4.3.3 beschrieben, auf Basis der Kookkurenzen auch ähnliche Begriffe zu den Suchbegriffen identifizieren. Soweit vorhanden, werden die einhundert ähnlichsten Begriffe zu jedem Suchbegriff gespeichert⁵³. Die tatsächlichen Ähnlichkeitswerte dieser Begriffe

⁵¹ vgl. Abschnitt 4.2

⁵² vgl. Abschnitt 4.3.2

⁵³ Im folgenden ist immer die Beschränkung auf diese einhundert ähnlichsten Begriffe gemeint, wenn von „ähnlichen Begriffen“ die Rede ist.

können betragsmäßig gering sein, genutzt werden soll im Rahmen der Suchterweiterung jedoch lediglich die Tatsache, ob jeweils zwei Suchbegriffe gemeinsame ähnliche Begriffe haben. Im WiBA-Net Textkorpus existiert für Gruppen aus drei oder mehr Suchbegriffen meist maximal ein und sehr oft gar kein gemeinsamer ähnlicher Begriff. Deshalb werden lediglich die möglichen Zweiwort-Paarungen der eingegebenen Suchbegriffe untersucht und deren gemeinsame ähnliche Begriffe zur weiteren Verarbeitung gespeichert.

Die Grundlage des erweiterten Suchvektors bilden die Stammformen der eingegebenen Suchbegriffe. Diesen ursprünglichen Suchbegriffen werden weitere Suchbegriffe zugefügt, um den Recall der Suchanfrage, also den Anteil aller insgesamt im System vorhandenen relevanten Wissensseinheiten, die unter den zurückgelieferten Suchergebnissen sind, möglichst groß werden zu lassen. Die zugefügten Suchbegriffe sind

- alle Begriffe, die den Kombinationen aus je zwei Suchbegriffen gemeinsam ähnlich sind,
- die signifikanten Kookkurenzen, die allen eingegebenen Suchbegriffen gemeinsam sind und
- die jeweils zehn Kookkurenzen mit den höchsten Signifikanzwerten zu den ursprünglichen Suchbegriffen.

Begriffe, die mehrere der angegebenen Kriterien erfüllen, werden nur einmal in den Suchvektor aufgenommen.

Diese auf dem Kontext basierende Erweiterung des Suchvektors erlaubt auch fachlich nicht vorgebildeten Anwendern eine einfache Benutzung des Systems. Studierende müssen fachsprachliche Suchterme nicht vorab kennen, um Informationen über die dahinter stehenden Inhalte und Zusammenhänge zu bekommen. Aus wenigen konkreten Suchbegriffen wird ein Vektor mit verwandten Begriffen erzeugt, der auch Überschneidungen der Kontexte der verschiedenen Suchbegriffe berücksichtigt.

Zu beachten ist, dass bei Suchbegriffen insbesondere geringer Häufigkeit auch Begriffe mit niedriger Trennschärfe im erweiterten Suchvektor auftreten können (z.B. allgemeinsprachliche Begriffe). Diese liefern, im Gegensatz zu Fachausdrücken, relativ viele irrelevante Seiten zurück. Ferner können im Suchvektor Fachausdrücke enthalten sein, die nur peripher etwas mit dem Thema der eigentlichen Suchanfrage zu tun haben und die ebenfalls irrelevante Seiten zurück liefern. Irrelevante Seiten werden anschließend durch die noch folgenden Schritte weitgehend entfernt.

Im Folgenden wird zwischen „Suchbegriffen“ und den „ursprünglichen Suchbegriffen“ unterschieden. Erstere umfassen den gesamten erweiterten Suchvektor, letztere nur die tatsächlich vom Nutzer eingegebenen Begriffe bzw. deren Stammformen. Die Unterscheidung wird benötigt, da ein Entscheidungskriterium bei der Relevanzbestimmung der gefundenen Seiten ist, ob die ursprüngli-

chen Suchbegriffe im Text oder in der Überschrift der betreffenden Seite enthalten sind. Ein weiteres Entscheidungskriterium beruht auf der Ähnlichkeit der Wissensseinheiten zu einem auf den ursprünglichen Suchbegriffen basierenden Vektor. Beide Entscheidungskriterien werden in Abschnitt 5.2.3 erläutert.

5.2.2 Suchergebnisse

Die erweiterten Suchterme werden als Suchvektor SV gespeichert. Der nächste Schritt ist das Auffinden aller Seiten, die einen, mehrere oder alle Begriffe des Vektors SV enthalten.

Dazu werden alle Wissensseinheiten, in denen einer oder mehrere der Terme des Suchvektors vorkommen, anhand der inversen Liste, die die Zuordnung der Terme zu den sie enthaltenden Dateien abbildet, identifiziert und gespeichert. Diese Ergebnismenge E_0 enthält auch Seiten, in denen die Suchbegriffe nur in abgeleiteter Form oder als Konstituent eines Kompositums vorkommen. Wegen der im Suchvektor nicht zu vermeidenden allgemeinsprachlichen Ausdrücke und ggf. vorhandener, nicht auf das Thema bezogener Fachausdrücke wird hierbei in der Regel eine große Menge an Seiten zurückgeliefert. Die Anzahl der zurückgelieferten Seiten wird bei häufigen Suchbegriffen noch erhöht⁵⁴. Diese weisen durchschnittlich mehr gemeinsame Kookkurenzen mit anderen Begriffen auf als seltene Terme und vergrößern folglich den erweiterten Suchvektor gegenüber diesen. Die Ergebnismenge zeichnet sich wegen ihres Umfangs durch einen nur geringen Precision-Wert bezüglich der Suchanfrage aus.

5.2.3 Eingangsgrößen für die Relevanzberechnung

Für jedes der gefundenen Dokumente werden Ähnlichkeitsvergleiche mit verschiedenen, aus den Suchbegriffen abgeleiteten Vektoren durchgeführt. Hierzu gehört zunächst die Ähnlichkeit zwischen dem Dokument d_i und dem Suchvektor SV $\text{sim}_{SV}(d_i)$. Weitere wichtige Aussagen lassen sich durch zwei zusätzliche Ähnlichkeitsmaße gewinnen. Zur Berechnung des ersten Ähnlichkeitsmaßes wird für jedes Dokument festgehalten, durch welche Suchbegriffe im Suchvektor es gefunden wurde. Für diese Begriffe werden jetzt die signifikanten Kookkurenzen aus der Datenbank ausgelesen. Diese Kookkurenzen bilden den globalen Kontext der Suchbegriffe, mit denen die Seite gefunden wurde. Sie werden – ohne die jeweiligen Suchbegriffe – zum Vergleichsvektor FV („findende“ Suchbegriffe) zusammengestellt. Der zweite Vergleichsvektor OV („Original“-Suchbegriffe) basiert auf den ursprünglichen Suchbegriffen. In diesen Vektor werden alle ursprünglichen Suchbegriffe sowie ihre jeweils 25 Kookkurenzen mit den höchsten Signifikanzwerten aufgenom-

⁵⁴ Bei einer Suche nach „Stahl“ und „Beton“ werden z.B. auf diese Weise 1302 von 2040 vorhandenen Dateien gefunden (entspricht ca. 63,8 % aller Seiten). Von diesen Seiten ist nur ein kleiner Teil für den zu erstellenden Lehrpfad relevant.

men. Mit diesen Vektoren werden die Ähnlichkeiten $\text{sim}_{\text{FV}}(d_i)$ und $\text{sim}_{\text{OV}}(d_i)$ berechnet. Alle Ähnlichkeiten werden mit dem Kosinusmaß nach Formel (2.1) bestimmt, die Werte geben also den Kosinus des Winkels zwischen den betrachteten Vektoren an.

Hierbei wird ein Sachverhalt beobachtet, der die Vergleichbarkeit der Ergebnisse erschwert. Die Wissenseinheiten des WiBA-Net, aufgebaut nach dem Prinzip, in einer Wissenseinheit nur einen Inhalt abzubilden, variieren sehr stark in ihrer Länge. Im Gegensatz zur Ähnlichkeitsberechnung zwischen zwei Dokumenten soll bei der Ähnlichkeitsberechnung zwischen einem (Such-)Vektor und einem Dokument die größtmögliche Ähnlichkeit ($\text{sim}_{\text{XV}}(d_i) = 1,0$) nicht erst dann vorliegen, wenn die beiden zu vergleichenden Elemente identisch sind, sondern schon, wenn alle Einträge des Suchvektors in dem Dokument enthalten sind. Dies führt aber praktisch dazu, dass für einen fiktiven, im Vergleich zu den Dokumentenvektoren relativ niedrigdimensionalen Suchvektor SV, dessen Einträge in zwei unterschiedlich langen Dokumenten jeweils vollständig vorhanden sind, unterschiedliche Ähnlichkeiten $\text{sim}_{\text{SV}}(d_i)$ und $\text{sim}_{\text{SV}}(d_j)$ berechnet werden. Derselbe Unterschied zwischen den beiden Ähnlichkeitswerten tritt auch auf, wenn nicht alle Begriffe des fiktiven Suchvektors in den beiden Dokumenten vorhanden sind, sondern jeweils nur eine gleichgroße Teilmenge dieser Begriffe. Dieser Unterschied wird im Kontext der vorliegenden Arbeit als nicht sinnvoll betrachtet, da die Relevanz eines Dokuments für eine Suchanfrage bzw. für die Generierung eines Lehrpfads nicht von dessen Länge abhängen soll.

Aus diesem Grund werden die zwischen Suchvektoren und Wissenseinheiten berechneten Ähnlichkeiten normiert. Ausgangspunkt ist die Überlegung, dass die Ähnlichkeit zu 1,0 werden soll, wenn alle Einträge des Suchvektors im Dokument enthalten sind. Also wird zunächst die Ähnlichkeit $\text{sim}_{\text{max}}(d_i)$ für ein (fiktives) Vorkommen aller Suchterme im betreffenden Dokument bestimmt. Dies geschieht nach Formel (2.1). Im Vergleichsdokument mehrfach auftretende Worte werden nur einmal berücksichtigt, deshalb lässt sich für den Zähler auch der Betrag des Suchvektors einsetzen. Dadurch wird die Berechnung beschleunigt. Alle tatsächlich berechneten Ähnlichkeiten $\text{sim}_{\text{vorh}}(d_i)$ werden dann auf diesen Maximalwert bezogen. Es gilt für die korrigierten Ähnlichkeiten $\text{sim}_{\text{kor}}(d_i)$:

$$\text{sim}_{\text{kor}}(d_i) = \frac{\text{sim}_{\text{vorh}}(d_i)}{\text{sim}_{\text{max}}(d_i)} \quad (5.2)$$

Auf dieser Basis werden die Ähnlichkeiten des Dokuments d_i zu den Vektoren SV, FV und OV berechnet und gespeichert. Für die zu korrigierende, tatsächliche Ähnlichkeit gilt jeweils:

$$\text{sim}_{\text{vorh}, \text{SV}}(D_i) = \cos(\alpha_{\text{SV}, D_i}) = \frac{\text{SV} \cdot D_i}{|\text{SV}| \cdot |D_i|} \quad (5.3)$$

$$\text{sim}_{\text{vorh}, \text{FV}}(D_i) = \cos(\alpha_{\text{FV}, D_i}) = \frac{\text{FV} \cdot D_i}{|\text{FV}| \cdot |D_i|} \quad (5.4)$$

$$\text{sim}_{\text{vorh}, \text{OV}}(D_i) = \cos(\alpha_{\text{OV}, D_i}) = \frac{\text{OV} \cdot D_i}{|\text{OV}| \cdot |D_i|} \quad (5.5)$$

Die maximalen Ähnlichkeiten sim_{max} können nur für einfache Häufigkeiten sinnvoll berechnet werden, also nicht, falls mehrfach auftretende Worte in den Dokumentvektoren mehrfach berücksichtigt werden. Dies wäre z.B. bei der Berechnung der Ähnlichkeit zweier Dokumente der Fall. Die Ähnlichkeit zwischen zwei Dokumenten kann folglich nicht normiert werden.

Zusätzlich zu den Ähnlichkeitswerten werden für die Bestimmung der Relevanz der Wissenseinheiten noch Informationen über das Vorkommen der ursprünglichen Suchbegriffe in den Texten und in den Überschriften der betrachteten Dokumente benötigt. Der Anteil der ursprünglichen Suchbegriffe, der sich in der Überschrift einer Seite findet, wird mit A_{UE} bezeichnet, der Anteil, der sich im Fließtext findet mit A_{FT} . Es gilt:

$$A_{\text{UE}}(d_i) = \frac{\text{Anzahl der ursprünglichen Suchbegriffe in der Überschrift}}{\text{Anzahl der ursprünglichen Suchbegriffe}} \quad (5.6)$$

$$A_{\text{FT}}(d_i) = \frac{\text{Anzahl der ursprünglichen Suchbegriffe im Fließtext}}{\text{Anzahl der ursprünglichen Suchbegriffe}} \quad (5.7)$$

Die fünf Größen $\text{sim}_{\text{SV}}(d_i)$, $\text{sim}_{\text{FV}}(d_i)$, $\text{sim}_{\text{OV}}(d_i)$, $A_{\text{UE}}(d_i)$ und $A_{\text{FT}}(d_i)$ sind Eingangsgrößen für die im folgenden Abschnitt erläuterte Relevanzberechnung.

5.3 Relevanz der Seiten

5.3.1 Zugehörigkeit von Wissenseinheiten zum Lehrpfad

Für alle mit einer Suchanfrage gefundenen Seiten muss entschieden werden, ob sie Bestandteil des zu generierenden Lehrpfads werden sollen. Dies geschieht durch die Berechnung von Relevanzwerten für jede Seite. Für die möglichen Relevanzwerte wird eine Skala von 0 bis 100 definiert. Die Seiten mit den höchsten Relevanzwerten werden dann in der nächsten Stufe weiterverarbeitet und zu einem Lehrpfad strukturiert.

Die Bestimmung der Relevanzwerte für ein Dokument d_i geschieht mit Hilfe einer Fuzzy-Logic-Berechnung. Dabei wird aus den fünf Eingangswerten

- Ähnlichkeit des Dokuments mit dem Suchvektor,

- Ähnlichkeit des Dokuments mit den Kookkurenzen der Suchbegriffe, die in dem Dokument vorhanden sind,
- Ähnlichkeit des Dokuments mit den ursprünglichen Suchbegriffen und deren signifikantesten Kookkurenzen,
- Anteil der ursprünglichen Suchbegriffe, die im Fließtext des Dokuments enthalten sind und
- Anteil der ursprünglichen Suchbegriffe, die in der Überschrift des Dokuments enthalten sind

abgeleitet, wie hoch die Relevanz des Dokuments bezüglich des Themas der Suchanfrage ist. Dabei wird nicht versucht, aus einigen Eingangsgrößen die Reihenfolge und aus anderen die Struktur des Lehrpfads zu bestimmen. Stattdessen wird durch Berücksichtigung aller Eingangsgrößen auf nur eine Ausgangsgröße geschlossen. Diese stellt dann den Relevanzwert R_i für das Dokument d_i bezüglich der Suchanfrage dar. Dieser Wert ist nur für die Entscheidung wichtig, ob eine Wissensseinheit in den zu erstellenden Lehrpfad aufgenommen wird. Die Gliederung des Lehrpfads und die Reihenfolge der Wissensseinheiten im Lehrpfad werden in einem weiteren Schritt bestimmt.

Die fünf Eingangsgrößen haben unterschiedliche Einflüsse auf die Relevanz eines Dokuments bzw. die Zugehörigkeit einer Wissensseinheit zum Lehrpfad:

- Die Ähnlichkeit sim_{SV} mit dem Suchvektor stellt die Übereinstimmung des Inhalts einer Wissensseinheit mit der Gesamthematik der Suche dar. Die gefundenen Ähnlichkeiten sind relativ niedrig, da in der Regel nur Teile eines Suchvektors in einzelnen Dokumenten gefunden werden. Die Berücksichtigung dieses Ähnlichkeitswerts erlaubt zu prüfen, ob ein Dokument zu einem ausreichenden Grad der vorgegebenen Thematik entspricht.
- Die Ähnlichkeit sim_{FV} zu den signifikanten Kookkurenzen der Suchbegriffe, die tatsächlich auf der Seite gefunden werden, gibt Aufschluss darüber, ob das Thema der untersuchten Seite tatsächlich dem des jeweiligen Suchterms entspricht. Ist der Suchterm außerhalb seines eigentlichen Kontextes auf der Seite enthalten, so liefert diese Ähnlichkeitsberechnung niedrige Werte.
- Die Ähnlichkeit sim_{OV} zu den ursprünglichen Suchbegriffen stellt einen Test dar, ob die ursprünglichen Suchbegriffe und ihre Kontexte in einer Wissensseinheit hinreichend abgedeckt sind. Ist der Ähnlichkeitswert niedrig, so enthält das Dokument vornehmlich durch die Erweiterung hinzu gekommene Suchbegriffe und deren Kontexte und ist deshalb möglicherweise themenfremd.
- Der Anteil A_{FT} der ursprünglichen Suchbegriffe, die im Fließtext der Seite vorkommen, stellt einen konkreteren Bezug zum Thema der eigentlichen Suche dar, als es der Vektor der Originalsuchbegriffe kann, da dieser auch die jeweiligen Kookkurenzen enthält. Da nicht in allen Wissensseinheiten, die der Thematik der Suchanfrage entsprechen, auch die Original-

suchbegriffe enthalten sein müssen, stellt A_{FT} ein die Ähnlichkeit sim_{OV} zu den ursprünglichen Suchbegriffen ergänzendes Kriterium dar.

- Der Anteil A_{UE} der ursprünglichen Suchbegriffe, die in der Überschrift des Dokuments vorkommen, kann die Wichtigkeit einer Wissensseinheit für das Thema der Suchanfrage ergänzend charakterisieren. Dieses Kriterium ist bei einem hohen Wert jedoch bedeutend aussagekräftiger als bei einem niedrigen Wert: Eine Überschrift, die alle Original-Suchbegriffe enthält, bezeichnet mit hoher Wahrscheinlichkeit eine relevante Seite. Eine Wissensseinheit mit einer Überschrift, die keinen dieser Suchbegriffe enthält, muss deswegen jedoch keineswegs irrelevant sein.

Für die Berechnung werden zunächst die scharfen Eingangswerte (konkrete Ähnlichkeitswerte und Anteile) in unscharfe Größen transformiert („Fuzzifizierung“). Aus diesen unscharfen Eingangsgrößen wird dann mittels Inferenzregeln auf unscharfe Ausgangsgrößen geschlossen. Diese unscharfen Ausgangsgrößen werden abschließend in einen scharfen Ausgangswert transformiert („Defuzzifizierung“). Dieser scharfe Ausgangswert ist der zu bestimmende Relevanzwert.

5.3.2 Fuzzifizierung

Die Eingangswerte für die Ähnlichkeiten sim_{xv} und für die Anteile A_v der einzelnen Wissensseinheiten liegen zunächst als konkrete, scharfe Werte vor. Sie müssen für die Fuzzy-Berechnung erst in unscharfe Daten transformiert werden. Für alle Ähnlichkeiten soll eine Einordnung in die Kategorien „geringe Ähnlichkeit“, „mittlere Ähnlichkeit“ und „hohe Ähnlichkeit“ erfolgen. Um diese Einteilung vornehmen zu können, muss zunächst definiert werden, was die jeweiligen Ähnlichkeitsklassen bedeuten, bzw. wie die entsprechende Zugehörigkeitsfunktion verlaufen soll.

Die Aussage, dass es sich bei einer bestimmten Ähnlichkeit, ausgedrückt als Kosinus des Winkels der verglichenen Vektoren, also im Intervall $[0,1]$ liegend, um eine hohe oder eine geringe Ähnlichkeit handelt, kann leider nicht verallgemeinert werden. Da die Ähnlichkeiten $sim_{sv}(d_i)$, $sim_{fv}(d_i)$ und $sim_{ov}(d_i)$ in unterschiedlichen Wertebereichen liegen und auch die jeweils gleichen Ähnlichkeitswerte bei unterschiedlichen Suchbegriffen nicht genau im selben Wertebereich liegen, muss für jede Suchanfrage eine individuelle Zugehörigkeitsfunktion erstellt werden⁵⁵. Zur Veranschaulichung zeigt Abbildung 5.3 eine typische Verteilung der Ähnlichkeitswerte. In dieser Abbildung ist exemplarisch die Verteilung der Ähnlichkeiten für die Suchanfrage „Stahl – Beton – Korrosion“ dargestellt. Insgesamt ergeben sich 1190 Treffer. Eingetragen sind die Mittelwerte der Ähnlichkeitsklassen, die die Breite 0,05 haben.

⁵⁵ Da man den prinzipiellen Verlauf der Zugehörigkeitsfunktion beibehalten kann, müssen lediglich die Parameter verändert werden.

Die Verteilungen der drei Ähnlichkeiten lassen sich durch die folgenden Überlegungen erklären. Man erkennt bei den Ähnlichkeiten zum Suchvektor (SV) hauptsächlich niedrige Werte (Maximalwert: 0,46; Mittelwert: 0,062) und eine geringe Spreizung. Das liegt daran, dass nur wenige der Suchbegriffe auf den Seiten tatsächlich vorhanden sind. Eine „hohe“ Ähnlichkeit liegt in der Regel schon bei drei oder vier gefundenen Suchbegriffen vor. Die Kookkurenzen der auf der betrachteten Seite vorhandenen Suchbegriffe weisen im Schnitt höhere Ähnlichkeitswerte (Maximalwert: 0,69; Mittelwert 0,199) mit dem betrachteten Dokument auf und sind weiter verteilt (FV). Dies liegt hauptsächlich an den höherdimensionalen Vergleichsvektoren FV. Diese sind wegen der in der Regel hohen Anzahl an Kookkurenzen der findenden Suchbegriffe nicht sehr themenspezifisch und liefern deswegen auch viele höhere Ähnlichkeitswerte. Die Ähnlichkeiten der ursprünglichen Suchbegriffe und ihrer Kookkurenzen (OV) ähneln in ihrer Verteilung denen des Suchvektors SV, weisen jedoch hohe Werte in den niedrigen Ähnlichkeitsklassen auf (Maximalwert: 0,29; Mittelwert: 0,043). Diese Verteilung ist damit zu erklären, dass der Vektor OV einen geringen Umfang hat und sehr themenspezifisch ist. Deswegen weisen nur wenige Seiten höhere Ähnlichkeiten mit dem Vektor OV auf.

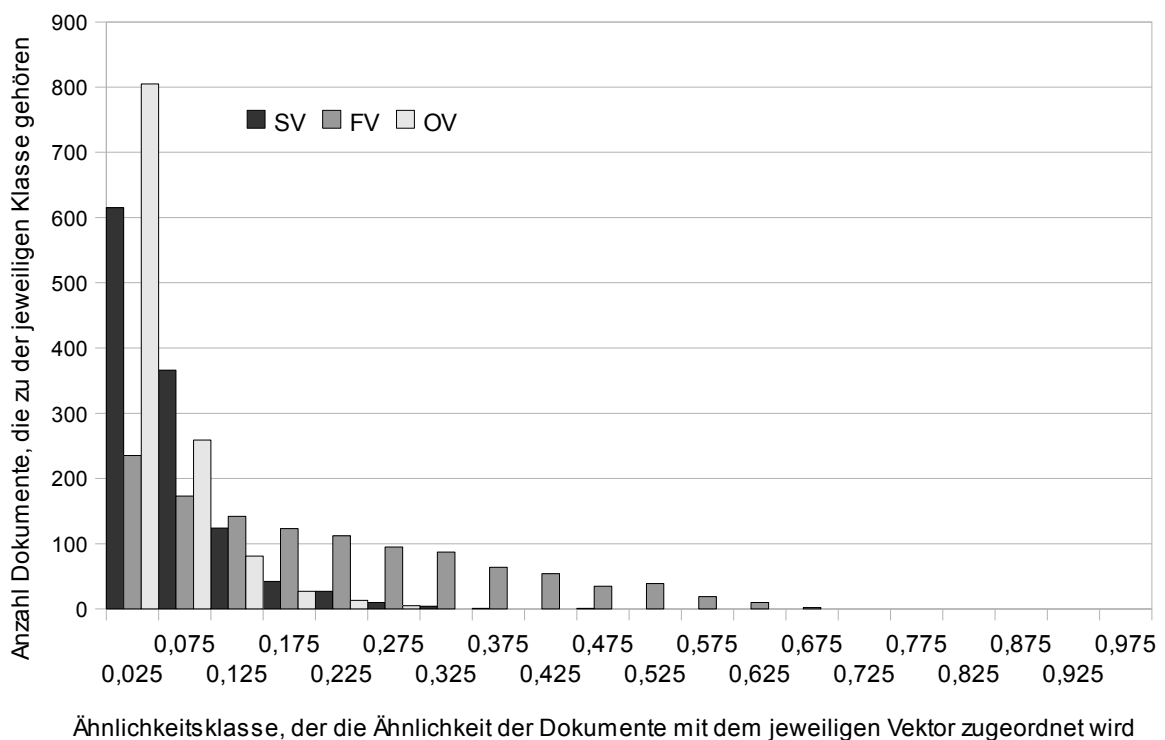


Abbildung 5.3: Typische Verteilung der Ähnlichkeiten der bei einer Suche gefundenen Dokumente

Um die Zugehörigkeitsfunktionen zu bestimmen, werden für alle Ähnlichkeitsverteilungen der Mittelwert und der Wert $W_{99,5}$ berechnet. $W_{99,5}$ ist der Wert, unterhalb dessen 99,5 % aller auftretenden

Werte liegen, er entspricht dem 99,5%-Quantil $Q_{0,995}$. Als Zugehörigkeitsfunktionen werden Dreiecks- und Trapezfunktionen verwendet (vgl. Abbildung 5.4). Der Bereich mittlerer Ähnlichkeit beginnt beim Mittelwert der Verteilung und endet beim berechneten Wert $W_{99,5}$ (oberer Grenzwert).

Der obere Grenzwert muss sehr hoch gewählt werden, da die Dokumente, die die oberen Rangplätze belegen, mit ausreichender Trennschärfe voneinander abgegrenzt werden sollen. Die Anzahl der Dokumente, deren Ähnlichkeitswerte oberhalb des oberen Grenzwertes liegen, muss deutlich kleiner sein als die geplante Länge des Lehrpfads, da sonst zu viele Dokumente Zugehörigkeitswerte von genau 1,0 zu den unscharfen Mengen mit der Eigenschaft „hohe Ähnlichkeit“ erhalten, unabhängig vom tatsächlichen Ähnlichkeitswert. Liegen die Werte unterhalb des oberen Grenzwertes, also auf dem ansteigenden Bereich des Graphen, so ergeben sich Werte ungleich 1,0, die vom Ähnlichkeitswert abhängig sind und damit für jedes Dokument nahezu „individuell“. Außerdem ergeben sich so auch Zugehörigkeiten zur unscharfen Menge „mittlere Ähnlichkeit“. Als Konsequenz weist die Ergebnismenge Relevanzwerte auf, die sich in der Regel eindeutig voneinander unterscheiden, wenn auch teilweise nur im Nachkommabereich. Eine eindeutige Reihenfolge der Relevanzen kann so aber erstellt werden. Eine Ausnahme bilden Seiten, die den höchsten möglichen Relevanzwert erhalten. Hier treten nicht selten mehrere gleiche Werte in einem Lehrpfad auf.

Die Zugehörigkeitsgrade μ_i werden als Wert der Zugehörigkeitsfunktion an der Stelle des Eingangswertes bestimmt. Da sich der Bereich mittlerer Ähnlichkeit entweder mit dem geringen oder dem hohen Ähnlichkeitsbereich überschneidet, sind je Eingangswert i.d.R. zwei Zugehörigkeitsgrade zu bestimmen. Für Abbildung 5.4 ergeben sich z.B. für ein Dokument, dessen Ähnlichkeit zum Suchvektor SV dem Wert $\text{sim}_{SV} = (5MW + W_{99,5}) / 6$ entspricht, Zugehörigkeitsgrade $\mu_{SV,gering}(\text{sim}_{SV}) = 2/3$ und $\mu_{SV,mittel}(\text{sim}_{SV}) = 1/3$. Die Zugehörigkeitswerte für einen Eingangswert summieren sich immer zu eins.

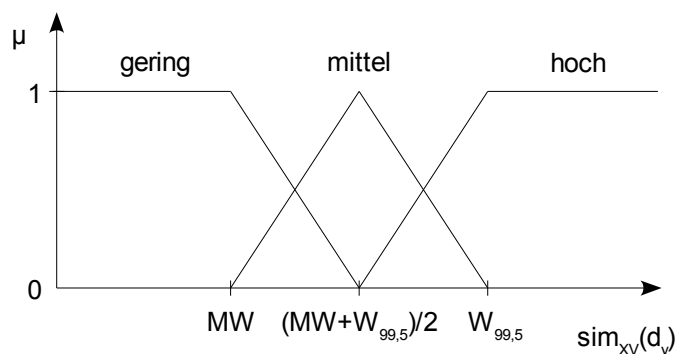


Abbildung 5.4: Zugehörigkeitsfunktion zu den unscharfen Mengen der Ähnlichkeit in Abhängigkeit von Mittelwert MW und 99,5%-Quantil der Verteilung

Für jedes Dokument wird für jede Ähnlichkeit (sim_{SV} , sim_{FV} , sim_{OV}) der Zugehörigkeitsgrad zu jeder der unscharfen Mengen (μ_{gering} , μ_{mittel} , μ_{hoch}) bestimmt. Man erhält also jeweils neun Zugehörigkeitsgrade.

Für die Anteile A_{FT} und A_{UE} der ursprünglichen Suchbegriffe, die Bestandteil von Fließtext bzw. Überschrift sind, werden ebenfalls Zugehörigkeitsfunktionen definiert. Diese sind jedoch nicht von einer Verteilung abhängig. Als Beginn des Bereichs mittlerer Ähnlichkeit wird ein Anteil von 20 % angesetzt (sodass bei einer angenommenen Anzahl von fünf Suchbegriffen einer im betrachteten Fließtext bzw. in der Überschrift vorhanden ist). Der Bereich endet bei 60 % (sodass ein Anteil von zwei Dritteln der Suchbegriffe in Fließtext bzw. Überschrift im Bereich hoher Ähnlichkeit liegt). Es ergibt sich die Zugehörigkeitsfunktion nach Abbildung 5.5.

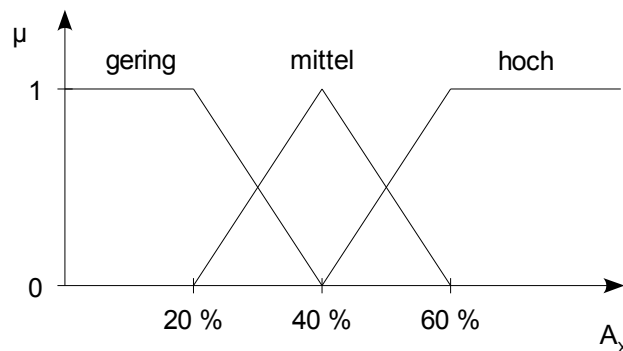


Abbildung 5.5: Zugehörigkeitsfunktion zu den unscharfen Mengen der Anteile der Suchbegriffe, der in Fließtext bzw. Überschrift vorhanden ist

Wiederum wird für jedes Dokument für beide Suchbegriffsanteile (FT, UE) der Zugehörigkeitsgrad zu allen unscharfen Mengen bestimmt. Es ergeben sich je Dokument sechs weitere Zugehörigkeitsgrade.

Eine Zusammenstellung aller fuzzyfizzierten Eingangswerte als Zugehörigkeitsgrade μ für ein betrachtetes Dokument d_i enthält Tabelle 5.4.

Tabelle 5.4: Zugehörigkeitsgrade zu den unscharfen Mengen

Merkmal	Ähnlichkeit / Anteil		
	gering	mittel	hoch
$\text{sim}_{\text{SV}}(d_i)$	$\mu_{\text{SV,gering}}(d_i)$	$\mu_{\text{SV,mittel}}(d_i)$	$\mu_{\text{SV,hoch}}(d_i)$
$\text{sim}_{\text{FV}}(d_i)$	$\mu_{\text{FV,gering}}(d_i)$	$\mu_{\text{FV,mittel}}(d_i)$	$\mu_{\text{FV,hoch}}(d_i)$
$\text{sim}_{\text{OV}}(d_i)$	$\mu_{\text{OV,gering}}(d_i)$	$\mu_{\text{OV,mittel}}(d_i)$	$\mu_{\text{OV,hoch}}(d_i)$
$A_{\text{UE}}(d_i)$	$\mu_{\text{UE,gering}}(d_i)$	$\mu_{\text{UE,mittel}}(d_i)$	$\mu_{\text{UE,hoch}}(d_i)$
$A_{\text{FT}}(d_i)$	$\mu_{\text{FT,gering}}(d_i)$	$\mu_{\text{FT,mittel}}(d_i)$	$\mu_{\text{FT,hoch}}(d_i)$

5.3.3 Inferenz

Die unscharfen Mengen müssen jetzt miteinander verknüpft werden, sodass Ausgangswerte entstehen, die allerdings auch noch unscharf sind. Für diese Verknüpfung werden Verarbeitungsregeln aufgestellt und angewendet. Sie stellen Bedingungen auf, bei deren Erfüllung eine bestimmte Folgerung eintritt.

Auch für die unscharfen Ausgangsgrößen müssen Zugehörigkeitsfunktionen definiert werden. Dies geschieht wieder mittels Dreiecks- und Trapezfunktionen. Diesmal sind jedoch fünf Bereiche definiert, die in Abbildung 5.6 dargestellt sind. Es werden fünf Bereiche gewählt, um bei der Zuordnung der verschiedenen Kombinationen der Zugehörigkeitsgrade der Eingangsgrößen zu den Ausgangsgrößen mehr Variationsmöglichkeiten zu haben. Die Relevanz als scharfe Ausgangsgröße ist im Rahmen der vorliegenden Arbeit, wie bereits beschrieben, auf einer Skala von 0 bis 100 als relative, dimensionslose Größe definiert. Ein Relevanzwert einer bestimmten Größe sagt also nicht direkt etwas darüber aus, ob die betreffende Wissensseinheit tatsächlich in den Lehrpfad aufgenommen wird, sondern nur, wo sie sich in die Reihenfolge der Wissensseinheiten einordnet. Wie oben beschrieben sollen die Beträge der Relevanzen möglichst nicht identisch sein, sodass jeder Wissensseinheit ein eindeutiger Rangplatz zugewiesen werden kann. Außerdem lässt sich die Berechnung leichter nachvollziehen, wenn eine Änderung des Eingangswerts auch eine Änderung des Ausgangswerts zur Folge hat und dieser nicht auf den Maximalwert begrenzt ist. Voraussetzung dafür ist allerdings eine eindeutige, individuelle Kombination aus Eingangsgrößen. Diese ist bei fünf Eingangsgrößen, die auf unterschiedlichen Ähnlichkeiten beruhen, zu erwarten.

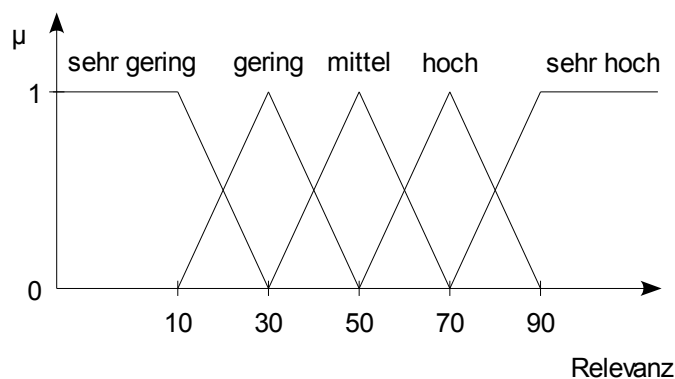


Abbildung 5.6: Zugehörigkeitsfunktionen der unscharfen Mengen der Ausgangsgrößen

Das Grundprinzip bei der Abarbeitung der Inferenzregeln ist, dass der Zugehörigkeitsgrad einer Schlussfolgerung dem der Prämisse entspricht. Werden mehrere Prämissen zu einer Schlussfolgerung verknüpft, so werden Operatoren zur Bestimmung des Zugehörigkeitsgrades der Schluss-

folgerung verwendet. Eine UND-Verknüpfung wird durch den Minimum-Operator abgebildet, eine ODER-Verknüpfung durch den Maximum-Operator.

Auf jede Wissensseinheit werden alle Inferenzregeln angewendet. Sobald eine Verknüpfung von Prämissen auf die Eingangsgrößen der Wissensseinheit zutrifft, wird die Schlussfolgerung berechnet. Die Schlussfolgerungen der vorliegenden Fuzzy-Berechnung ordnen die Wissensseinheiten den verschiedenen Kategorien der Relevanz⁵⁶ für den Lehrpfad zu. Wird eine Wissensseinheit durch mehrere Prämissen derselben Kategorie zugeordnet, so wird zwischen den jeweiligen Werten der Zugehörigkeit zu dieser Kategorie der Mittelwert berechnet.

Die für die Relevanzbestimmung verwendeten Inferenzregeln beruhen auf einem stufenweisen Entwicklungsprozess. Die erste Stufe war die Formalisierung offensichtlicher Zusammenhänge zwischen Kombinationen von Eingangsgrößen und den resultierenden Ausgangsgrößen als Inferenzregeln⁵⁷. Die zweite Stufe war die Ergänzung der erhaltenen Regeln durch Variation einzelner Eingangsgrößen und die Festlegung von Schlussfolgerungen zu den Ausgangsgrößen, die sich daraus ergeben sollten. Auf der Basis der so erweiterten Regeln konnten in der dritten Stufe erste Relevanzwerte für Wissensseinheiten bestimmt werden. Deren Auswertung wurde iterativ zu einer weiteren Modifizierung und Ergänzung der Regeln genutzt, bis die Ergebnisse zufriedenstellend waren. In der vierten Stufe wurden noch fehlende Inferenzregeln für bislang noch nicht berücksichtigte Kombinationen von Eingangsgrößen logisch ergänzt.

In Tabelle 5.5 sind die Inferenzregeln dargestellt, die zur Ermittlung der Ausgangsgrößen auf die unscharfen Eingangswerte für das jeweils betrachtete Dokument d_i angewendet werden.

Tabelle 5.5: Inferenzregeln für die Bestimmung der Zugehörigkeit von Wissensseinheiten zu einem Lehrpfad

Zeile	sim _{SV}	sim _{FV}	A _{UE}	g	g	g	m	m	m	h	h	h
			sim _{OV} \ A _{FT}	g	m	h	g	m	h	g	m	h
1	h	h	h	0	+	++	+	++	++	++	++	++
2	h	h	m	0	+	+	+	+	+	+	+	++
3	h	h	g	--	0	0	0	0	0	0	+	+
4	h	m	h	0	+	+	+	+	+	+	+	++
5	h	m	m	-	-	0	-	0	0	0	+	+
6	h	m	g	--	-	-	--	-	-	--	-	-
7	h	g	h	0	+	+	+	+	+	+	+	++
8	h	g	m	-	-	0	-	0	0	0	0	+
9	h	g	g	--	--	-	--	--	-	--	-	-

56 „sehr gering“, „gering“, „mittel“, „hoch“, „sehr hoch“

57 z.B. sehr hohe Zugehörigkeit zum Lehrpfad, wenn alle Eingangsgrößen den Wert „hoch“ haben

Zeile	sim _{SV}	sim _{FV}	A _{UE}	g	g	g	m	m	m	h	h	h
			sim _{OV} \ A _{FT}	g	m	h	g	m	h	g	m	h
10	m	h	h	0	+	+	+	+	+	+	++	++
11	m	h	m	0	0	0	0	0	+	+	+	+
12	m	h	g	--	-	-	--	0	0	--	0	+
13	m	m	h	0	0	0	0	0	0	0	+	++
14	m	m	m	--	-	0	0	0	0	0	0	+
15	m	m	g	--	-	-	--	-	-	--	-	-
16	m	g	h	-	-	-	0	0	0	0	+	++
17	m	g	m	--	--	-	0	0	0	0	0	0
18	m	g	g	--	--	--	--	--	-	--	-	-
19	g	h	h	0	+	+	+	+	+	+	++	++
20	g	h	m	--	--	-	-	-	0	0	+	+
21	g	h	g	--	--	--	--	-	0	--	-	+
22	g	m	h	-	-	0	0	0	0	0	+	++
23	g	m	m	--	--	-	-	-	-	0	0	+
24	g	m	g	--	--	--	--	--	--	--	--	--
25	g	g	h	-	-	-	0	0	0	0	+	+
26	g	g	m	--	--	--	--	-	-	--	-	0
27	g	g	g	--	--	--	--	--	--	--	--	--

Abhängig von der Zugehörigkeit der fünf Eingangsgrößen zu den Kategorien „hoch“ (h), „mittel“ (m) oder „gering“ (g) ist die jeweilige Relevanz einer Wissenseinheit, die die entsprechenden Eigenschaften aufweist, für den zu erstellenden Lehrpfad angegeben. Hierbei bedeuten

- „++“: sehr hohe Relevanz
- „+“: hohe Relevanz
- „0“: mittlere Relevanz
- „-“: geringe Relevanz
- „--“: sehr geringe Relevanz

Zu beachten ist, dass Wissenseinheiten häufig zwei Kategorien angehören. Im Folgenden sollen der Aufbau der Inferenzregeln und ihre Anwendung erläutert werden.

Die Inferenzregeln sind der Übersichtlichkeit halber in Tabellenform dargestellt. Die Zeilen stellen die verschiedenen Kombinationen der Ähnlichkeitsgrößen sim_{SV}, sim_{FV} und sim_{OV} dar, die Spalten

die Kombinationen der Suchbegriffsanteile A_{FT} und A_{UE} . Die Gruppierung zu jeweils neun Werten geschieht anhand jeweils gleicher Werte für sim_{FV} und A_{UE} und dient hauptsächlich der leichteren Lesbarkeit der Tabelle. Unterschiedliche Zugehörigkeiten innerhalb einer Neunergruppe folgen ausschließlich aus den Veränderungen der Eingangsgrößen sim_{OV} und A_{FT} . Ferner sind auch die Blöcke von Zeile 1 bis 9 (Block 1), von Zeile 10-18 (Block 2) und von Zeile 19-27 (Block 3) durch eine dickere Umrandung hervorgehoben. Grundsätzlich verlaufen innerhalb einer Neunergruppe die Zugehörigkeiten von links nach rechts und von unten nach oben monoton in der Richtung von geringeren zu höheren Zugehörigkeiten. Des Weiteren ist der Wert an einer bestimmten Position innerhalb einer Neunergruppe in der unterhalb folgenden Gruppe (oder innerhalb eines Blocks im unterhalb folgenden Block) immer größer oder gleich dem Wert in der unteren Gruppe (oder im unteren Block). Dies folgt aus der Art der gewählten Eingangsgrößen, da jede größer werdende Eingangsgröße auch eine potentiell größere Relevanz der entsprechenden Wissensseinheit nach sich zieht. Gegenläufige Elemente treten nicht auf.

Die Tabelle kann in Form von Verarbeitungsregeln abgearbeitet werden. So kann das „+“ in der siebten Spalte der zweiten Zeile umschrieben werden als

„WENN Ähnlichkeit zum Suchvektor hoch UND Ähnlichkeit zum Vektor der findenden Begriffe hoch UND Ähnlichkeit zu den ursprünglichen Suchbegriffen mittel UND Anteil der Suchbegriffe in der Überschrift mittel UND Anteil der Suchbegriffe im Fließtext gering, DANN Relevanz für den Lehrpfad hoch“,

für die Spalten 1 bis 6 in Zeile 13 lässt sich schreiben:

„WENN Ähnlichkeit zum Suchvektor mittel UND Ähnlichkeit zum Vektor der findenden Begriffe mittel UND Ähnlichkeit zu den ursprünglichen Suchbegriffen hoch UND (Anteil der Suchbegriffe in der Überschrift gering ODER Anteil der Suchbegriffe in der Überschrift mittel), DANN Relevanz für den Lehrpfad mittel“.

Auf dieser Grundlage wird der Einfluss der einzelnen Größen bewertet. Durch eine Abwägung zwischen den verschiedenen Einflüssen gelangt man dann zu den verwendeten Inferenzregeln. Es wird davon ausgegangen, dass der Einfluss der Suchbegriffsanteile A_{FT} und A_{UE} auf die Zugehörigkeit einen eher multiplikativen Charakter hat, der vorhandene Tendenzen der Zugehörigkeit verstärkt. Folglich verläuft die Steigung zwischen zwei benachbarten Relevanzwerten von links nach rechts in der Regel flacher als von unten nach oben. Die Suchbegriffsanteile A_{FT} und A_{UE} stellen durch die alleinige Berücksichtigung der Suchworte, ohne deren Kontext mit einzubeziehen, einen Übergang zur Booleschen Suche dar. Ihr gleichzeitiger Vor- und Nachteil ist die Nichtberücksichtigung des Kontextes: Ein hoher Wert bezieht sich präzise auf vorhandene Suchbegriffe, ein niedriger Wert muss jedoch nicht bedeuten, dass die Wissensseinheit nicht im Kontext eines der Suchbe-

griffe steht. Deshalb kann eine Wissensseinheit mit guten Ähnlichkeitswerten sim_{xv} auch für niedrige Suchbegriffsanteile A_{v} gut platziert werden, umgekehrt jedoch nicht. Man sieht dies z.B. an den Zeilen 1, 4, 7 und 10, die auch im Bereich der niedrigen Suchbegriffsanteilswerte keine geringen Zugehörigkeiten aufweisen. Demgegenüber findet sich bei niedrigen Ähnlichkeitswerten, z.B. in den Zeilen 15, 18, 24 und 27 auch für hohe Anteilswerte keine mittlere oder hohe Zugehörigkeit.

Der Wert sim_{sv} basiert auf dem am wenigsten homogenen Vergleichsvektor. SV ist der erweiterte Suchvektor und definiert damit das vollständige gesuchte Themenfeld. FV bezieht sich nur auf die in einer Wissensseinheit gefundenen Suchbegriffe, OV nur auf die ursprünglichen Suchbegriffe im engeren Sinn. Deshalb lassen sich die sim_{sv} -Werte als eine Art Grundrelevanz verstehen. Die drei Blöcke der Tabelle werden durch jeweils gleiche sim_{sv} -Werte gebildet. Man erkennt den analogen Aufbau der Zugehörigkeiten in diesen Blöcken, wenn auch auf unterschiedlichem Niveau.

In der ersten Zeile haben alle Ähnlichkeiten sehr hohe Werte, entsprechend sind die Relevanzbewertungen verteilt. Falls einer der Werte für A_{FT} oder A_{UE} nur einen mittleren Wert erreicht, wird die Zugehörigkeit auf „hoch“ abgeschwächt, sind beide entsprechenden Werte nur gering, so erhält die Zugehörigkeit einen mittleren Wert. Die zweite und dritte Zeile haben einen ähnlichen Verlauf auf jeweils niedrigerem Zugehörigkeitsniveau, da der sim_{ov} -Wert, die Ähnlichkeit mit den ursprünglichen Suchbegriffen, sinkt. Da dieser Wert einen großen Einfluss auf die Gesamt-Relevanz hat, beginnt die oberste Zeile einer Neunergruppe immer mit hohen Werten. Hierbei werden auch höhere Werte als in der untersten Zeile der vorangegangenen Neunergruppe verwendet. Die Werte für sim_{fv} oder auch sim_{sv} sind in diesem Fall zwar gesunken, jedoch kann dieser Effekt durch den wichtigen sim_{ov} -Wert wieder ausgeglichen werden. Dennoch sinken auch die Werte jeweils gleicher Zeilen einer Neunergruppe gegenüber der darüber befindlichen Neunergruppe monoton.

Die Ähnlichkeit mit dem Suchvektor sim_{sv} ist wichtig zur Identifikation von Seiten, die keinen direkten Bezug zu den ursprünglichen Suchtermen haben, sondern z.B. Synonyme verwenden. Der Wert sim_{sv} hat auch bei den Seiten, die lediglich einen ähnlichen Kontext wie ein Suchbegriff haben, höhere Werte. Aus diesem Grund vergleicht man bei der Bestimmung des Ähnlichkeitswerts sim_{sv} eine Wissensseinheit jeweils mit dem Kontext der gesamten Suchanfrage, der durch den erweiterten Suchvektor SV dargestellt wird.

Durch die Ähnlichkeit sim_{fv} zum Vektor der Kookkurrenten der Suchbegriffe, die in der Seite enthalten sind, kann der Einfluss eventueller Ausreißer berücksichtigt werden. Mit Ausreißern werden Treffer, die eigentlich keine sind, bezeichnet, also Seiten, die die jeweiligen gefundenen Begriffe nicht als zentrales Thema enthalten⁵⁸. Der Einfluss von sim_{fv} ist der geringste, kann jedoch im Ein-

⁵⁸ Dies kann z.B. der Fall sein, wenn auf der entsprechenden Seite ein zentral behandelter Begriff mit anderen Begriffen verglichen wird.

zelfall wichtig sein. Die Werte der Zeilen 1-9, 10-18 und 19-27 fallen deshalb jeweils weniger stark ab als die entsprechenden Werte der Zeilen 1-3, 10-12, 19-21 etc.⁵⁹.

Ziel der Inferenz ist es, dass Seiten, die nicht zum Lehrpfad passen, niedrige Relevanzwerte erhalten, deshalb im Rang sinken und nicht in den Lehrpfad gelangen. Umgekehrt sollen passende Seiten natürlich hohe Relevanzwerte erhalten. Es ist dabei zu beachten, dass in der Fuzzyfizierung jede Wissensseinheit für jede Einflussgröße mehreren Kategorien zugeordnet wird, wenn sie nicht im Sonderfall zu 100% zu einer Kategorie gehört. Im vorliegenden Fall wird jede Wissensseinheit für alle Einflussgrößen zwei Kategorien zugeordnet, nämlich entweder „gering“ und „mittel“ oder „mittel“ und „hoch“. Die betreffende Wissensseinheit gehört dabei den beiden Kategorien jeweils zu unterschiedlichen Anteilen an. Diese Anteile summieren sich zu 1,0. Gehört eine Wissensseinheit zu einem gewissen Anteil der hohen Kategorie einer Einflussgröße an, so hat dies einen erhöhenden Einfluss auf den Relevanzwert. Durch die Zugehörigkeit zur mittleren Kategorie ergibt sich je nach Einflussgröße ein neutraler, ein positiver oder ein negativer Einfluss auf die Relevanz. Eine Zugehörigkeit zum geringen Anteil wirkt negativ auf den Relevanzwert.

Die Inferenzregeln gelten für alle Wissensseinheiten. Einer Wissensseinheit wird also für jede Regel, die auf sie passt, ein Zugehörigkeitsgrad zugeordnet. Im Allgemeinen wird die Wissensseinheit dabei den Ergebniskategorien auch mehrfach zugeordnet: z.B. können sich aus der Abarbeitung zweier Inferenzregeln zwei unterschiedliche Zugehörigkeitsgrade zur Ergebnismenge „sehr hoch“ ergeben. Liegen mehrere Zugehörigkeitsgrade zu einer Kategorie für eine Wissensseinheit vor, so wird aus diesen als unscharfer Ergebniswert der arithmetische Mittelwert bestimmt.

Zum Abschluss der Inferenz müssen die berechneten Zugehörigkeitsgrade noch auf die Zugehörigkeitsfunktionen der Ausgangsgrößen übertragen werden [TRAEGER 1994]. Dies geschieht nach der in Abschnitt 2.2.3.3 beschriebenen Methode der Teilschwerpunkte. Dazu wird die Fläche unter jeder Zugehörigkeitsfunktion aus Abbildung 5.6 mit dem entsprechenden Zugehörigkeitsgrad multipliziert. Es ergibt sich der Gesamtschwerpunkt aller Teilflächen zur Weiterverarbeitung in der Defuzzyfizierung. In Abbildung 5.7 ist die Teilfläche für sehr geringe Relevanz mit einem durch Inferenz bestimmten Zugehörigkeitsgrad von 0,5 dargestellt.

59 Hiervon gibt es eine Ausnahme: Ist der sim_{SV} -Wert einer Wissensseinheit in Kombination mit deren A_{FT} -Wert hoch, so kann daraus gefolgert werden, dass Original-Suchbegriffe in der Wissensseinheit enthalten sind. Deren Kontext, ausgedrückt durch den Vektor FV , ergibt ebenfalls einen hohen Ähnlichkeitswert. Die entsprechenden Relevanzwerte werden deshalb höher angesetzt, als es sonst der Fall wäre, wenn der sim_{SV} -Wert mittel oder gering ist. Zum Vergleich dienen die Zeilen 12 und 21, in denen der sim_{SV} -Wert mittel bzw. gering ist, der sim_{FV} -Wert jedoch hoch, und umgekehrt die Zeilen 6 und 9, in denen der sim_{FV} -Wert gering oder mittel und der sim_{SV} -Wert hoch ist. Diese Überlegung ist nur theoretischer Natur, da die Kombination von hohem sim_{SV} -Wert und hohem A_{FT} -Wert in aller Regel auch einen hohen sim_{OV} -Wert ergeben wird. Die erwähnten Zeilen, in denen man den Effekt berücksichtigen muss, haben also nur geringe praktische Bedeutung. An dieser Stelle können die oben erwähnten Zeilen 15, 18, 24 und 27, bei denen auch für hohe Ähnlichkeitswerte keine mittlere oder hohe Zugehörigkeit zugeordnet wird, betrachtet werden. In der Reihe der aufgezählten Zeilen fehlt die Nummer 21, eben weil hier aus den beschriebenen Gründen doch mittlere und hohe Zugehörigkeiten zugeordnet werden.

Die Vorgehensweise bei der Inferenz soll im Folgenden an einem konkreten Beispiel dargestellt werden. Eine Wissensseinheit habe folgende Zugehörigkeitswerte zu den Eingangsgrößen:

- $\mu_{SV,mittel} = 0,15$; $\mu_{SV,hoch} = 0,85$
- $\mu_{FV,mittel} = 0,54$; $\mu_{FV,hoch} = 0,46$
- $\mu_{OV,hoch} = 1,0$
- $A_{FT,hoch} = 1,0$
- $A_{UE,gering} = 0,33$; $A_{UE,mittel} = 0,67$

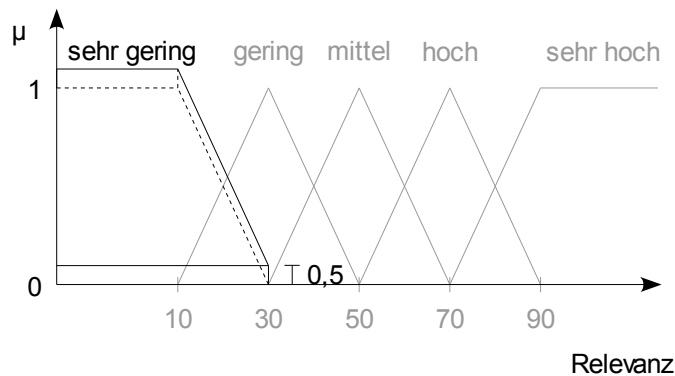


Abbildung 5.7: Beispiel zur Inferenz mit der Methode der Teilschwerpunkte

Es ergeben sich aus diesen Zugehörigkeitsgraden acht mit den Inferenzregeln zu untersuchende Kombinationsmöglichkeiten:

1. $\mu_{SV,hoch}$; $\mu_{FV,hoch}$; $\mu_{OV,hoch}$; $A_{FT,hoch}$; $A_{UE,gering}$
2. $\mu_{SV,hoch}$; $\mu_{FV,hoch}$; $\mu_{OV,hoch}$; $A_{FT,hoch}$; $A_{UE,mittel}$
3. $\mu_{SV,hoch}$; $\mu_{FV,mittel}$; $\mu_{OV,hoch}$; $A_{FT,hoch}$; $A_{UE,gering}$
4. $\mu_{SV,hoch}$; $\mu_{FV,mittel}$; $\mu_{OV,hoch}$; $A_{FT,hoch}$; $A_{UE,mittel}$
5. $\mu_{SV,mittel}$; $\mu_{FV,hoch}$; $\mu_{OV,hoch}$; $A_{FT,hoch}$; $A_{UE,gering}$
6. $\mu_{SV,mittel}$; $\mu_{FV,hoch}$; $\mu_{OV,hoch}$; $A_{FT,hoch}$; $A_{UE,mittel}$
7. $\mu_{SV,mittel}$; $\mu_{FV,mittel}$; $\mu_{OV,hoch}$; $A_{FT,hoch}$; $A_{UE,gering}$
8. $\mu_{SV,mittel}$; $\mu_{FV,mittel}$; $\mu_{OV,hoch}$; $A_{FT,hoch}$; $A_{UE,mittel}$

Die Kombinationen 1 und 2 haben als Schlussfolgerung nach den Inferenzregeln die Relevanz „sehr hoch“, die Kombinationen 3-6 die Relevanz „hoch“ und die Kombinationen 7 und 8 die Relevanz „mittel“. Es ergeben sich folgende Werte für die Kombinationen:

1. $\mu_{Relevanz, \text{ sehr hoch}} = \min(\mu_{SV,hoch}; \mu_{FV,hoch}; \mu_{OV,hoch}; A_{FT,hoch}; A_{UE,gering}) = 0,33$
2. $\mu_{Relevanz, \text{ sehr hoch}} = \min(\mu_{SV,hoch}; \mu_{FV,hoch}; \mu_{OV,hoch}; A_{FT,hoch}; A_{UE,mittel}) = 0,46$
3. $\mu_{Relevanz, \text{ hoch}} = \min(\mu_{SV,hoch}; \mu_{FV,mittel}; \mu_{OV,hoch}; A_{FT,hoch}; A_{UE,gering}) = 0,33$

4. $\mu_{\text{Relevanz, hoch}} = \min(\mu_{\text{SV, hoch}}, \mu_{\text{FV, mittel}}, \mu_{\text{OV, hoch}}, A_{\text{FT, hoch}}, A_{\text{UE, mittel}}) = 0,54$
5. $\mu_{\text{Relevanz, hoch}} = \min(\mu_{\text{SV, mittel}}, \mu_{\text{FV, hoch}}, \mu_{\text{OV, hoch}}, A_{\text{FT, hoch}}, A_{\text{UE, gering}}) = 0,15$
6. $\mu_{\text{Relevanz, hoch}} = \min(\mu_{\text{SV, mittel}}, \mu_{\text{FV, hoch}}, \mu_{\text{OV, hoch}}, A_{\text{FT, hoch}}, A_{\text{UE, mittel}}) = 0,15$
7. $\mu_{\text{Relevanz, mittel}} = \min(\mu_{\text{SV, mittel}}, \mu_{\text{FV, mittel}}, \mu_{\text{OV, hoch}}, A_{\text{FT, hoch}}, A_{\text{UE, gering}}) = 0,15$
8. $\mu_{\text{Relevanz, mittel}} = \min(\mu_{\text{SV, mittel}}, \mu_{\text{FV, mittel}}, \mu_{\text{OV, hoch}}, A_{\text{FT, hoch}}, A_{\text{UE, mittel}}) = 0,15$

Für jede Kategorie der Relevanz wird der Mittelwert bestimmt, man erhält als Abschluss der Inferenz folgende Zugehörigkeitsgrade der Ausgangswerte:

1. $\mu_{\text{Relevanz, sehr hoch}} = (0,33 + 0,46) / 2 = 0,40$
2. $\mu_{\text{Relevanz, hoch}} = (0,33 + 0,54 + 0,15 + 0,15) / 4 = 0,29$
3. $\mu_{\text{Relevanz, mittel}} = (0,15 + 0,15) / 2 = 0,15$

5.3.4 Defuzzifizierung

Abschließend muss aus den durch die Inferenz gewonnenen unscharfen Ergebnisdaten wieder ein scharfer Ausgangswert berechnet werden. Dieser stellt eine direkte Bewertung der Relevanz dar. Je höher der defuzzifizierte Ausgabewert für eine Wissenseinheit also ist, desto relevanter ist sie für die entsprechende Suchanfrage. Durch die große Anzahl von Inferenzregeln und die verteilten Eingangswerte ergeben sich i.d.R. unterschiedliche Relevanzwerte für die einzelnen Wissenseinheiten. Es liegt nur ein geringer Anteil, nämlich 0,5 %, der Seiten im jeweils obersten, konstanten Bereich der Zugehörigkeitsfunktionen für „sehr hohe“ Ähnlichkeit. Wenn eine Wissenseinheit mit allen Eingangswerten in diesem Bereich angesiedelt ist, so ergeben sich dieselben Ausgangswerte auch für leicht voneinander differierende Eingangsgrößen. Sobald jedoch zumindest in einer Kategorie Werte im linearen Bereich einer Zugehörigkeitsfunktion dazukommen, werden diese in aller Regel maßgeblich, da durch den Minimum-Operator die jeweils kleinsten Werte für die Bestimmung einer unscharfen Ausgangsgröße genutzt werden⁶⁰. Die Trennschärfe zur Auswahl der relevantesten Seiten ist also gegeben. Die Wahrscheinlichkeit, dass mehrere Seiten denselben Relevanzwert aufweisen, ist vor allem im kritischen Bereich um den Rangplatz L herum, wo sich entscheidet, ob eine Seite in den Lehrpfad aufgenommen wird oder nicht, gering.

Der größte mögliche Relevanzwert liegt im Schwerpunkt der Fläche unter der Zugehörigkeitsfunktion für „sehr hohe“ Relevanz und beträgt damit 89,17. Da dieser Wert nicht übertroffen werden kann, die Relevanz also auch für sehr relevante Seiten bei diesem Wert gedeckelt ist, tritt dieser

⁶⁰ vgl. Abschnitt 2.2.3.3

Wert gelegentlich bei mehreren Wissenseinheiten im selben Lehrpfad auf. Dies stellt kein Problem dar, da i.d.R. die Lehrpfadlänge deutlich größer ist als der Umfang des mehrfachen Auftretens⁶¹.

Die Defuzzifizierung erfolgt nach der Methode der Teilschwerpunkte. Dazu müssen die Flächeninhalte unterhalb der Zugehörigkeitsfunktion und deren Schwerpunkte bestimmt werden. Diese Berechnung braucht nur einmal durchgeführt zu werden, denn die errechneten Größen sind unabhängig von den tatsächlich vorhandenen unscharfen Ergebniswerten. Die Flächeninhalte werden dann mit den jeweils errechneten Zugehörigkeitsgraden multipliziert. Man erhält so gleichsam eine Gewichtung der jeweiligen Teilflächen. Die errechneten Produkte $\mu_i \cdot A_i$ werden im Schwerpunkt der jeweiligen Teilfläche angesetzt. Daraus wird dann der Gesamtschwerpunkt berechnet. Die x-Koordinate dieses Gesamtschwerpunktes ist gleich dem Ausgabewert.

Für den x-Wert des Schwerpunktes x_R und damit die Größe der Relevanz R ⁶² gilt in Anlehnung an TRAEGER 1994 bei den vorhandenen fünf Teilmengen mit den Teilschwerpunkten x_i , den Teilflächen A_i und den Zugehörigkeitsgraden μ_i :

$$\text{Relevanz } R = x_R = \frac{\sum_{i=1}^5 x_i \cdot \mu_i \cdot A_i}{\sum_{i=1}^5 \mu_i \cdot A_i} \quad (5.8)$$

Mit dem zuvor bestimmten Sollwert der Lehrpfadlänge L können jetzt die relevantesten Wissenseinheiten als Grundlage für die Weiterverarbeitung ausgewählt werden. Diese werden in der letzten Stufe zu einem Lehrpfad zusammengestellt. Jedoch können durch die Berücksichtigung von Unterthemen ggf. noch einige Wissenseinheiten gekürzt oder ergänzt werden.

5.3.5 Berücksichtigung von Unterthemen

Die Menge der durch die Relevanzberechnung ermittelten Wissenseinheiten wird vor der eigentlichen Lehrpfaderstellung noch angepasst. Dabei können nach festgelegten Kriterien sowohl Seiten entfernt als auch hinzugefügt werden, der tatsächliche Umfang des erstellten Lehrpfads kann also vom Sollwert der Lehrpfadlänge L abweichen. Die Ergänzung oder Entfernung von Seiten geschieht auf der Grundlage von Unterthemen, die aus den vom Nutzer eingegebenen Suchbegriffen abgeleitet werden können.

Von der Suchanfrage eines Nutzers wird angenommen, dass sie den Wunsch nach Informationen zu einem die Suchbegriffe umfassenden bzw. übergreifenden Thema ausdrückt. Die Relevanzwer-

61 vgl. hierzu die Relevanzwerte R in Anhang J

62 Exemplarische Relevanzwerte für einen Lehrpfad sind in Abschnitt 6.2 angegeben.

te der Wissensseinheiten, die auf die in den vorhergehenden Abschnitten beschriebene Weise bestimmt werden, beziehen sich auf dieses übergreifende Thema.

Neben der durch die Suchanfrage definierten Gesamthematik existieren zusätzlich Unterthemen, die nur einen Teilbereich der Anfrage umfassen. Dies kann Grundlagen- oder Spezialwissen sein, das zum Verständnis oder zur Vertiefung der Inhalte der Gesamthematik nützlich oder sogar notwendig ist. Beispielsweise ist es für einen Lehrpfad zum Thema des Einflusses von Beton auf die Korrosion von Stahl notwendig, auch die Voraussetzungen zu erläutern, die überhaupt gegeben sein müssen, damit Stahl korrodieren kann.

Auf die vorliegende Problematik übertragen, kann man die Unterthemen als durch Kombinationen von Suchbegriffen vorgegeben betrachten. Sind mehr als zwei Suchbegriffe vorhanden, so lassen sich diese paarweise kombinieren. Es entstehen Begriffspaare, die je ein Unterthema definieren. Deshalb wird davon ausgegangen, dass die von je zwei Suchworten aufgespannten Themenfelder für das Verständnis eines aus diesen und anderen Suchbegriffen bestehenden Gesamthemas wichtig sind. Wissensseinheiten, die intensiv auf den Zusammenhang zwischen jeweils zwei Suchbegriffen eingehen, jedoch wegen eines oder mehrerer fehlender Suchbegriffe einen zu geringen Relevanzwert erreichen, um noch in den Lehrpfad aufgenommen zu werden, müssen grundsätzlich bei der Lehrpfaderstellung berücksichtigt werden.

Dies ist direkt bei der Relevanzberechnung auf Fuzzy-Basis nicht möglich, da spätestens ab vier Stichworten mit sechs möglichen Suchbegriffspaaren die Eingangsgrößen zu zahlreich werden, um noch sinnvolle Inferenzregeln aufstellen zu können. Die Unterthemen werden also gesondert nach der Relevanzberechnung berücksichtigt. Alle Wissensseinheiten, die potentiell in den Lehrpfad aufgenommen werden, stammen aus dem Bereich hoher Relevanzwerte. Deshalb ist sichergestellt, dass durch die nachträgliche Ergänzung keine völlig irrelevanten Wissensseinheiten in den Lehrpfad aufgenommen werden.

Untersucht werden durch Begriffspaare definierte Themenfelder. Sind nur zwei Suchbegriffe vorhanden, werden die Begriffe einzeln als Unterthema untersucht. Ebenfalls werden die einzelnen Begriffe betrachtet, wenn ein Begriffspaar keine gemeinsamen ähnlichen Begriffe aufweist. Das jeweilige, sich auf den Kontext des Begriffspaares bzw. Begriffs beziehende Themenfeld wird auf der Basis der ähnlichen Begriffe gebildet. Bei Begriffspaaren besteht der Vergleichsvektor für die Ähnlichkeitsberechnung aus den beiden Begriffen gemeinsamen ähnlichen Begriffen sowie deren jeweils zehn höchstsignifikanten Konkurrenten. Den Vergleichsvektor von Einzelbegriffen bilden dessen ähnliche Begriffe.

Normalerweise soll der Umfang des Lehrpfads, im Wesentlichen aus didaktischen Gründen, maximal dem berechneten Sollwert L entsprechen. Deshalb kann die Länge des Lehrpfads nicht belie-

big vergrößert werden. Der Umfang ΔL der an einen Lehrpfad maximal anzufügenden Seiten wird in Abhängigkeit vom Sollwert L festgelegt. Er lässt sich Tabelle 5.6 entnehmen. Die angegebenen Werte sind jedoch variabel, denn die Aufnahme neuer Seiten wird an der Stelle des steilsten Abfalls der Ähnlichkeitswerte der zum Anhängen zur Verfügung stehenden Seiten in einer 2 Seiten breiten Umgebung um ΔL beendet.

Untersucht werden die Wissensseinheiten, die sich auf den 20 Rangplätzen befinden, die den L relevantesten Seiten folgen. Die Anzahl der Unterthemen betrage k . Zunächst werden für jede der untersuchten Seiten Ähnlichkeitsberechnungen mit den Vergleichsvektoren der einzelnen Begriffspaare bzw. Begriffe durchgeführt. Dabei werden die normierten Ähnlichkeiten $sim_{UT,n}$ der Seite mit den einzelnen Unterthemen (UT) bestimmt.

Tabelle 5.6: Maximale Anzahl von Zusatzseiten zu Unterthemen in Abhängigkeit vom Sollwert der Lehrpfadlänge L

Sollwert L	Zusatzseiten ΔL
< 11	2
< 19	3
< 25	4
25	5

Es erfolgt danach eine Bewertung der untersuchten Wissensseinheiten hinsichtlich ihrer Ähnlichkeit mit den k verschiedenen Unterthemen. Zunächst werden die untersuchten Wissensseinheiten für jedes Unterthema n in eine Reihenfolge gebracht, die den absteigend geordneten Ähnlichkeiten $sim_{UT,n}$ entspricht. Der Ähnlichkeitswert einer Seite wird dann durch deren jeweiligen Rangplatz R dividiert. Aus Gründen der Vereinfachung wird nicht die Ähnlichkeit jedes Begriffspaars bzw. jedes Begriffs mit den Wissensseinheiten einzeln untersucht, sondern die Ähnlichkeiten werden über alle k Begriffspaare bzw. Begriffe gemittelt. Es gilt:

$$sim_{UT}(d_i) = \frac{\sum_{n=1}^k \frac{sim_{UT,n}(d_i)}{R_n}}{k} \quad (5.9)$$

Diese gemittelten Ähnlichkeitswerte werden wiederum absteigend sortiert, sodass man eine Liste der für die Unterthemen relevantesten Wissensseinheiten erhält. Diese wurden durch die Rangplätze auf den k Listen gewichtet. In einer Umgebung von 2 Rangplätzen um den Rangplatz ΔL wird jetzt nach der größten Ähnlichkeitsdifferenz zwischen zwei aufeinander folgenden Werten gesucht. Alle Wissensseinheiten, die in der Liste über diesem Maximum liegen und gleichzeitig einen Ähn-

lichkeitswert von mindestens 0,10 aufweisen, werden in den Lehrpfad aufgenommen, darunter liegende Wissensseinheiten nur, wenn sie einen Ähnlichkeitswert von mindestens 0,15⁶³ aufweisen.

Gegenüber dieser Ergänzung des Lehrpfads können auf der Basis des Einflusses von Unterthemen auch Seiten aus dem Lehrpfad entfernt werden. Dies ist bei Wissensseinheiten möglich, die insgesamt nur geringe Ähnlichkeitswerte mit den Unterthemen aufweisen. Auf diese Weise können rechnerisch relevante Seiten, die zum Hauptthema der Suche jedoch nur wenig Bezug haben, erkannt und ausgeschlossen werden. Es ist nicht vollständig zu vermeiden, dass Wissensseinheiten, die nur für einen der gesuchten Begriffe sehr relevant sind, deren Inhalt aber nicht zu den anderen Suchbegriffen passt, in den Lehrpfad aufgenommen werden. Dies geschieht u.a. bei Kohyponymen von Suchbegriffen⁶⁴: Einer der Suchbegriffe kommt in der nicht passenden Wissensseinheit in einem relevanten Kontext vor, die Wissensseinheit behandelt aber insgesamt ein Thema, das ein Kohyponym zu einem der anderen Suchbegriffe darstellt. Dieses Kohyponym hat einen ähnlichen Kontext wie der eigentliche Suchbegriff und die Ähnlichkeitsberechnung für sim_{SV} liefert deshalb keinen niedrigen Wert. Ein Beispiel ist die Suche nach „Anwendung“, „Nadelholz“ und „Laubholz“. Hierbei wird auch die Seite „Übersicht häufiger Kunststoff-Anwendungen im Bauwesen“ anfänglich in den Lehrpfad aufgenommen. Der gemeinsame relevante Begriff ist „Anwendung“, „Kunststoff“ als Baustoff stellt aber ein Kohyponym zu „Laubholz“ bzw. „Nadelholz“ dar. Eine solche Seite kann jedoch durch einen niedrigen Ähnlichkeitswert sim_{UT} zu den Unterthemen nachträglich ausgeschlossen werden, wie es in diesem Beispiel auch der Fall ist: Hier hat die angeführte Seite nur einen geringen Ähnlichkeitswert zum Kontext des Suchwortpaares „Laubholz – Nadelholz“.

Um entsprechend aus dem Lehrpfad auszuschließende Seiten zu finden, werden analog zum Vorgehen bei der Ergänzung eines Lehrpfads Ähnlichkeiten zwischen den Vergleichsvektoren der Unterthemen und den Wissensseinheiten bestimmt, jedoch diesmal zu den Wissensseinheiten, die bereits im Lehrpfad sind, sich also auf den ersten L Plätzen befinden. Die Berechnungen für das Ergänzen und das Entfernen von Wissensseinheiten werden dabei vor den tatsächlichen Änderungen vorgenommen, die Menge der L Wissensseinheiten im Lehrpfad ist also beim Ergänzen und beim Entfernen gleich. Für die einzelnen Ähnlichkeitswerte gilt ebenfalls Formel (5.9). Die Wissensseinheiten werden diesmal aufsteigend nach der Ähnlichkeit sortiert. Es werden jetzt für die ersten sechs – also die niedrigsten – Ähnlichkeitswerte die Differenzen zum jeweils nächsthöheren Wert und der prozentuale Unterschied auf der Basis des größeren Wertes bestimmt. Aus den Differenzen wird ein durchschnittlicher Abfall der Ähnlichkeitswerte berechnet. Es werden vom niedrigsten

63 Dies stellt die Aufnahme von Wissensseinheiten mit sehr hohen Ähnlichkeiten ungeachtet der ermittelten zusätzlichen Seitenzahl ΔL sicher. Es ist zu beachten, dass der reale Ähnlichkeitswert durch das Teilen durch den Rangplatz und die Mittelung kleiner wird, daher wird auch der Vergleichswert von 0,15 niedrig angesetzt.

64 vgl. Abschnitt 6.2

Ähnlichkeitswert her jetzt alle Wissenseinheiten entfernt, deren Differenz zum nächsthöheren Wert größer als die 1,5-fache mittlere Differenz ist, oder deren prozentualer Unterschied zum nächsthöheren Wert größer als 15% ist. Das Entfernen wird abgebrochen, sobald eine Seite im Lehrpfad verbleibt, also nicht entfernt wurde, auch wenn danach die Ausschlusskriterien noch einmal erfüllt sein sollten. Es wird ebenfalls spätestens nach der sechsten Wissenseinheit abgebrochen.

Die gekürzten bzw. die um zusätzliche Seiten ergänzten relevanten Wissenseinheiten stellen jetzt die Ergebnismenge E_1 dar.

5.3.6 Weiterführende Seiten

Als weiterführende Seiten können den einzelnen, in einem Lehrpfad enthaltenen Seiten thematisch ähnliche Wissenseinheiten zugeordnet werden. Dabei wird auch der berechnete Relevanzwert R der jeweils betrachteten Seite berücksichtigt: Je relevanter eine Seite für den Lehrpfad ist, desto sinnvoller ist die Bereitstellung von weiterführenden Seiten zu dieser Seite. Wenig relevante Lehrpfadseiten haben demnach eine geringere Chance, dass ihnen eine weiterführende Seite zugeordnet wird.

Zunächst wird für jede Seite d_{LP} in dem Lehrpfad anhand der gespeicherten Ähnlichkeitswerte überprüft, ob zu dieser Seite ähnliche Wissenseinheiten existieren. Als Grenzwert $sim_{dw}(d_{LP})$ für die Ähnlichkeit, ab der eine Wissenseinheit d_w als weiterführende Seite gilt, wird der Wert 0,25 festgelegt. Um die Relevanz der betrachteten Lehrpfadseite zu berücksichtigen, wird dieser Wert durch die auf die maximal mögliche Relevanz 89,17 bezogene Relevanz R geteilt. Wenn folglich gilt

$$sim_{dw}(d_{LP}) > 0,25 \cdot \frac{89,17}{R} \quad (5.10)$$

dann wird d_w zu einer weiterführenden Seite. Jede Seite wird einem Lehrpfad nur einmal als weiterführende Seite zugeordnet. Seiten, die bereits im Lehrpfad enthalten sind, können nicht gleichzeitig weiterführende Seiten zu anderen Lehrpfadseiten sein. Die weiterführenden Seiten sind nicht Bestandteil des erstellten Lehrpfads. Es wird lediglich auf sie verwiesen, um eine Vertiefung der Inhalte der sie referenzierenden Seiten zu ermöglichen. Deshalb werden sie nicht in der dem Lerner angezeigten Navigations-Struktur des Lehrpfads dargestellt, sondern als Link direkt in die Wissenseinheit eingebunden, mit der sie verknüpft sind. Die weiterführenden Seiten werden in den folgenden Abschnitten, die sich mit der Strukturierung der Lehrpfade befassen, nicht weiter berücksichtigt.

5.4 Anordnung der Lehrpfadseiten

5.4.1 Vorüberlegungen

Die Präsentation der für eine Suchanfrage relevanten Seiten in einer didaktisch sinnvollen Gruppierung, Reihenfolge und Gliederung ist eine der zentralen Aufgaben des Algorithmus. Dabei ist bereits die Definition des Ausdrucks „didaktisch sinnvoll“ nicht-trivial. Zunächst ist die angemessene Didaktik vom Lerner und dessen Lernkontext abhängig. Des Weiteren existieren unterschiedliche Lehrmeinungen, wie einem Lerner in einer bestimmten Situation bestimmtes Wissen zu vermitteln ist. Schließlich wäre es selbst dann noch problematisch, wenn man den aus Lerner- und Lehrersicht am besten geeigneten didaktischen Ablauf gefunden hätte, die einzelnen zu vermittelnden Inhalte den Schritten dieses didaktischen Ablaufs zuzuordnen.

Aus diesem Grund ist es angezeigt, zunächst eine oder mehrere einfache und möglichst allgemein anerkannte Arten der didaktischen Vorgehensweise festzulegen. Diese müssen als Randbedingung auch von einem Algorithmus nach logischen Regeln zusammengestellt werden können. Die Kriterien, wie eine Seite in einem bestimmten Kontext zu sortieren ist, müssen aus dem Inhalt der Seite selbst bzw. den auf den Inhalten des Textkorpus basierenden textstatistischen Informationen entnommen werden. Zu bestimmen sind nacheinander

- die Gruppierung der Wissensseinheiten in Clustern,
- die Reihenfolge dieser Cluster und
- für jedes Cluster die Hauptseite und die Nebenseiten.

Eine Möglichkeit der Herausbildung einer Lehrpfadstruktur aus den relevanten Suchergebnissen ist die so genannte Clusterbildung der gefundenen Dokumente. Diese kann nach LEWANDOWSKI 2005 z.B. zur Strukturierung von Suchmaschinenergebnissen benutzt werden, damit der Nutzer die für ihn relevante Information leichter auswählen kann. Das Clustering eignet sich jedoch ebenfalls für die vorliegende Problematik, nämlich die Bündelung von Wissensseinheiten, um diese dem Nutzer sequentiell und vollständig anzuzeigen.

Beim Clustering werden Dokumente – bzw. im vorliegenden Fall Wissensseinheiten – um verschiedene zentrale Punkte gruppiert. Das Single-Pass-Cluster-Verfahren [FERBER 2003] prüft z.B. für alle Dokumente nacheinander den Abstand zum „Zentroid“ bzw. Schwerpunkt der Dokumente eines vorhandenen Clusters. Ist der Abstand klein genug, wird das Dokument zum betreffenden Cluster hinzugefügt, ist er größer als ein bestimmter Schwellwert, so wird ein neuer Cluster angelegt. Ist beim ersten Dokument noch kein Cluster vorhanden, so wird ebenfalls ein neuer Cluster angelegt. Im vorliegenden Fall wird eine Modifizierung dieses Verfahrens angewendet.

Zu beachten ist der Unterschied des Clustering zur Klassifizierung, bei der Dokumente nach bestimmten Kriterien zuvor festgelegten Kategorien zugefügt werden. Beim Clustering liegen vorab keine Themen fest, denen die Dokumente zugeordnet werden sollen. Sie ergeben sich stattdessen implizit während der Berechnung aus der Zusammenstellung ähnlicher Seiten.

Nach dem Clustering wird die Reihenfolge der so erhaltenen Wissenscluster festgelegt. Außerdem werden die Cluster intern gegliedert, d.h. es werden eine Hauptseite und ggf. Nebenseiten bestimmt.

5.4.2 Clustering der Wissensseinheiten

Die automatisch generierten Lehrpfade sollen dem Nutzer in einer didaktisch sinnvollen Weise präsentiert werden, um dadurch die Lernmotivation und letztlich den Lernerfolg zu steigern. Betrachtet wurde in den vorigen Abschnitten bereits die Bestimmung einer didaktisch sinnvollen Menge von relevanten Wissensseinheiten. Es ist jetzt noch mit Gruppierung, Reihenfolge und Gliederung die Struktur des Lehrpfads festzulegen.

Um ähnliche Inhalte in einem engen Kontext von Basisinformationen und Vertiefungswissen darstellen zu können, werden die Wissensseinheiten zu Themenblöcken, so genannten Clustern, zusammengefasst. Bei der Bildung von Clustern wird vor allem auf die inhaltliche Nähe der in einem Cluster zusammengefassten Informationen Wert gelegt. Daneben sollen die erstellten Cluster aber auch didaktischen Kriterien genügen. Diese drücken sich hierbei in der Forderung nach einem möglichst einheitlichen Umfang der Cluster aus. Auf diese Weise kann ein Lerner den noch zurückzulegenden Lernweg besser abschätzen. Ist ein Cluster kleiner, als er es nach inhaltlichen Gesichtspunkten sein müsste, so kann ein Lerner daraus fälschlicherweise auf eine höhere Spezialisierung der Themen der enthaltenen Seiten schließen. Die tatsächlichen Ähnlichkeiten zwischen den Wissensseinheiten werden bei der Festlegung der Clusterumfänge gleichwohl berücksichtigt.

Das angewendete Verfahren des Clustering beruht auf dem Single-Pass-Cluster-Verfahren [FERBER 2003]. Im angewendeten Fall wird der auf den Dokumentvektoren beruhende Schwerpunktsvektor der Seiten eines Clusters bestimmt und die Ähnlichkeiten zwischen diesem Schwerpunktsvektor und den Seiten, die an das Cluster angefügt werden sollen, berechnet. Damit, analog zum o.a. Verfahren, die Überschreitung eines Grenzwerts überprüft werden kann, wird nicht die Ähnlichkeit $\cos(\alpha)$ betrachtet – die größer wird, je ähnlicher sich die Dokumente sind –, sondern die Differenz $1 - \cos(\alpha)$. Diese wird im folgenden als Clusterdifferenzmaß bezeichnet. Einander ähnliche Dokumente weisen ein geringeres Clusterdifferenzmaß auf als einander unähnliche. Liegt das Clusterdifferenzmaß unterhalb eines festgelegten Grenzwertes, so wird die Seite dem Cluster zugefügt (vgl. Abbildung 2.3). Der Schwerpunktsvektor eines Clusters wird nach dem Hinzufügen

einer Seite neu berechnet. Die nächste Seite wird dann mit diesem neuen Schwerpunktsvektor verglichen. Angefügt wird jeweils die dem Cluster ähnlichste Seite, danach wird der Schwerpunktsvektor neu berechnet. Dies kann dazu führen, dass eine andere Seite, die vorher dem Clusterkern ähnlich genug gewesen wäre, nach der Neuberechnung des Schwerpunktes nicht mehr in das Cluster aufgenommen wird (vgl. Abbildung 5.8). Bei einem zu großen Clusterdifferenzmaß zum ähnlichsten Dokument wird ein neuer Cluster angelegt. Der erste Schwerpunkt dieses neuen Clusters ist der Dokumentenvektor der einzigen zugehörigen Seite, der nach dem Hinzufügen der zweiten Seite zu diesem Cluster ebenfalls neu bestimmt wird.

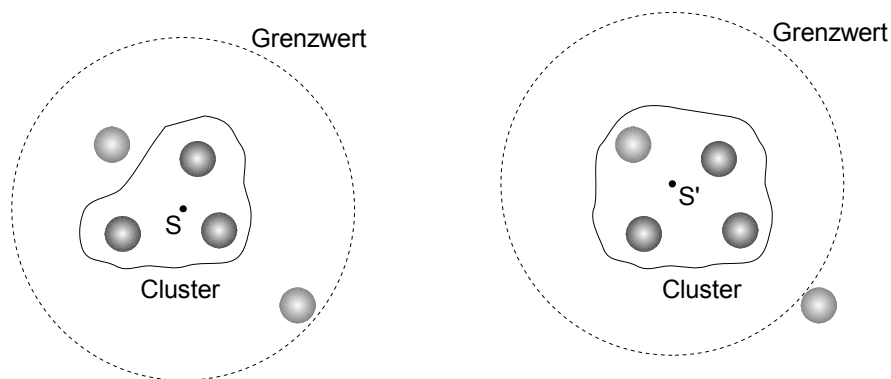


Abbildung 5.8: Cluster beim Anfügen neuer Seiten. Im Ursprungszustand liegen beide noch freien Wissenseinheiten unter dem Grenzwert des Clusterdifferenzmaßes um den Schwerpunkt S (links). Durch Zufügen der nächsten Wissenseinheit zum Cluster verlagert sich der Schwerpunkt zu S' , die zweite freie Wissenseinheit wird nicht mehr in den Cluster aufgenommen (rechts).

Das erste Cluster wird aus der Seite gebildet, die den höchsten Relevanzwert R aufweist. Es werden dann die Clusterdifferenzmaße zu dieser Seite bzw. diesem Clusterkern berechnet und für die nächste Seite wird nach dem oben beschriebenen Verfahren überprüft, ob sie in das Cluster aufgenommen werden soll.

Der Grenzwert G_c für die Zugehörigkeit zu einem Cluster ist spezifisch für den jeweiligen Lehrpfad. Er wird aus dem mittleren Clusterdifferenzmaß der Seiten untereinander vor dem Anlegen des ersten Clusters bestimmt. Aus allen Clusterdifferenzmaßen zwischen jeweils zwei Wissenseinheiten werden der Mittelwert m und die Standardabweichung s berechnet. Der Grenzwert wird dann mit $G_c = m - s$ festgelegt. Um die Bildung größerer Cluster gegenüber der Aneinanderreihung von Ein-Seiten-Clustern zu fördern, wird der Grenzwert bei neu angelegten, also bis auf die Startseite leeren Clustern, um 20 % erhöht. So steigt die Wahrscheinlichkeit, dass jeweils mindestens zwei Seiten zu einem Cluster verbunden werden. Eine Aneinanderreihung von Einzelseiten ist nur bei separaten Themen sinnvoll. Da, wie oben beschrieben, möglichst gleichgroße Cluster erzeugt wer-

den sollen, wird der Grenzwert zur Aufnahme von Wissensseinheiten in ein Cluster durch geeignete Parameter so modifiziert, dass möglichst gleichgroße Cluster entstehen.

Bei einem ab der zweiten Seite gleich bleibenden Grenzwert G_C tendieren die Seiten eines Lehrpfades dazu, sich im ersten Cluster zu gruppieren. Beim Aufnehmen einzelner Seiten verschiebt sich der Schwerpunkt eines Clusters, sodass nach und nach vom ursprünglichen Clusterkern zentrale, jedoch immer unähnlichere Seiten mit aufgenommen werden. Wenn von diesem „wandernden“ Cluster keine weiteren Seiten mehr aufgenommen werden, da diese zu weit entfernt sind, können die verbleibenden Seiten sich untereinander zu unähnlich sein, um noch zu Clustern gruppiert zu werden. Das Ergebnis ist ein großer Cluster und mehrere Einzelseiten als Cluster.

Die Agglomeration von Seiten im ersten Cluster kann dadurch vermindert werden, dass der Grenzwert G_C ab der dritten Seite um jeweils fünf Prozentpunkte pro Seite vermindert wird. Es gilt demnach für den Grenzwert $G_{C,n}$ eines Clusters vom Umfang n :

$$\begin{aligned} G_{C,1} &= 0,8 \cdot G_C & (\text{für } n < 2) \\ G_{C,2} &= G_C & (\text{für } n = 2) \\ G_{C,n} &= G_C + 0,05 \cdot n \cdot G_C & (\text{für } n > 2) \end{aligned} \quad (5.11)$$

Kann eine Seite mit dem geringsten Clusterdifferenzmaß – also der größten Ähnlichkeit – nicht aufgenommen werden, so scheiden alle anderen Seiten, die folglich eine geringere Ähnlichkeit aufweisen, ebenfalls aus. Muss ein neues Cluster angelegt werden, so wird die dem aktuellen Cluster unähnlichste Seite – also die mit dem größten Clusterdifferenzmaß – als Clusterkern ausgewählt. Auf diese Weise wird sichergestellt, dass die einzelnen Cluster sich thematisch möglichst weitgehend voneinander unterscheiden, um jeweils möglichst abgeschlossene Themengebiete zu umfassen. Außerdem ist auf diese Weise die Ähnlichkeit der Wissensseinheiten innerhalb der Cluster höher als zwischen den Clustern. Dies kommt der späteren hierarchischen Gliederung entgegen, da Nebenseiten zu den Hauptseiten ähnliche Seiten sind.

Im Anschluss an das Clustering müssen die Cluster in eine didaktisch sinnvolle Reihenfolge gebracht werden. Dies geschieht nach den in Abschnitt 5.4.3 vorgestellten Grundsätzen. Besteht ein Cluster aus mehr als einer Wissensseinheit, so müssen in diesem Kontext nicht die Wissensseinheiten sondern die Cluster sortiert werden. Hierzu werden deren Inhalte gemeinsam – gleichsam als eine Wissensseinheit – betrachtet.

5.4.3 Sortierung der Cluster

5.4.3.1 Ordnungsprinzipien

Aus den Wissensseinheiten an sich lassen sich einige didaktische Merkmale, die für die Festlegung einer Reihenfolge genutzt werden können, explizit auslesen. So ist z.B. das Wort „Einführung“ Bestandteil einiger Überschriften von Wissensseinheiten. Solche Merkmale weisen natürlich direkt auf eine geeignete Platzierung der entsprechenden Seite hin. Will man einen Lehrpfad nach dem Prinzip „vom Allgemeinen zum Besonderen“ aufbauen, so wäre die oben erwähnte Einführungsseite am besten für den Anfang dieses Lehrpfads geeignet. Da das entwickelte Programmsystem jedoch auch allgemeine Inhalte anderer Wissensdomänen ohne explizite Beschreibungen der Inhalte verwalten soll, muss auf die Nutzung dieser Informationen verzichtet werden. Ansonsten wäre bei Inhaltsbasen, die auf erläuternde Überschriften nach dem obigen Beispiel verzichten, keine Anwendung des Systems möglich.

Die Funktionalität des Systems beruht lediglich auf Rückschlüssen aus den linguistischen und textstatistischen Daten. Deshalb müssen einfache Regeln für die Erstellung didaktischer Strukturen gefunden werden, die mit hinreichender Sicherheit auch von einem Computerprogramm nachvollzogen werden können.

Der erste zu berücksichtigende didaktische Aspekt ist die geeignete Länge eines Lehrpfads, die nach Formel (5.1) bestimmt werden kann. Der zweite wichtige didaktische Aspekt ist die Reihenfolge der Wissensseinheiten bzw. der aus den Wissensseinheiten gebildeten Cluster. Die Wissensseinheiten stellen jeweils ein in sich abgeschlossenes Thema dar, bauen also nicht aufeinander auf und können deshalb grundsätzlich frei kombiniert werden. Gleiches gilt prinzipiell auch für die aus den Wissensseinheiten gebildeten Cluster. Trotzdem ist eine willkürliche Anordnung nicht erwünscht, da die Lerner die Wissensseinheiten bzw. Cluster sonst nicht oder nur schlecht in einen Kontext einordnen können. Die erstellten Cluster lassen sich nach verschiedenen Ordnungsprinzipien anordnen, dies sind nach TRIBELHORN 2008 beispielsweise:

- die Anordnung vom Bekannten zum Unbekannten,
- die Anordnung vom Konkreten zum Abstrakten,
- die Anordnung vom Einfachen zum Komplexen,
- die Anordnung vom Allgemeinen zum Speziellen und
- die Anordnung von der Ursache zur Wirkung.

Nahezu alle dieser Kategorien lassen sich für die Wissensseinheiten aufgrund der zur Verfügung stehenden textstatistischen Informationen programmbasiert zumindest grob einschätzen.

Ein nahe liegender Ansatz wäre, die Wissensseinheiten nach allen Kategorien zu bewerten und dann eine Anordnung von der Gruppe der Seiten, die hohe Werte für Allgemeinheit, konkrete Inhalte und Einfachheit haben, zu der Gruppe der Seiten, die hohe Werte für spezielle Inhalte, Abstraktion oder Komplexität haben, vorzunehmen. Dies scheidet jedoch aus, weil die Wissensseinheiten den Kategorien durchaus gegenläufig zugeordnet sein können. So kann eine allgemeine Seite gerade durch die umfassende Darstellung eines Sachverhalts sehr komplex sein. Eine spezielle Seite hingegen kann sehr einfach sein, wenn auf dieser z.B. nur ein Beispiel für einen Sachverhalt genannt wird.

Ebenfalls verworfen wird der Ansatz, dem Nutzer mehrere Gliederungsarten anzubieten, den Lehrpfad also in verschiedenen Varianten bereitzustellen, die den o.a. Ordnungsprinzipien entsprechen. Da die Gliederungen nach den Ordnungsprinzipien auf der Basis von Rückschlüssen aus textstatistischen bzw. linguistischen Informationen erfolgen, stellen sie kein exaktes Ergebnis dar. Vielmehr sind die erstellten Gliederungsvarianten als Ergebnisvorschläge unterschiedlicher Qualität zu betrachten. Von diesen ist der Vorschlag, bei dem es sich mit der höchsten Wahrscheinlichkeit um den besten handelt, auszuwählen.

Die Methode, mit der sich mit der höchsten Wahrscheinlichkeit ein didaktisch sinnvoller Lehrpfad zusammenstellen lässt, ist also, die Bewertung der Wissensseinheiten hinsichtlich aller Kategorienpaare durchzuführen und die Reihenfolge der Cluster im Lehrpfad dann nach der Anordnung mit der höchsten Unterscheidungskraft festzulegen. Ein Lerner hat so zwar nicht die Wahlmöglichkeit einer didaktischen Methode, es wird aber die jeweils beste mögliche, d.h. die am sichersten objektiv nachvollziehbare, didaktische Anordnung ausgewählt. Hochschullehrer haben ohnehin die Möglichkeit, die vorgeschlagene Reihenfolge nach ihren eigenen Vorstellungen zu verändern. Im Folgenden werden die einzelnen Ordnungsprinzipien hinsichtlich ihrer Anwendbarkeit im vorliegenden Fall betrachtet.

Am einfachsten zu realisieren wäre die Anordnung der Cluster vom Bekannten zum Unbekannten. Dem Lerner bekannte Seiten können leicht aufgezeichnet werden und diesbezüglich ausgewertet werden. Der konkrete Wissensstand des Lerners ist indessen zunächst unbekannt, wenn er nicht mit einer Lernerfolgskontrolle bestimmt wird. Jedoch kann nicht für alle potentiell einsetzbaren Inhalte eine solche Lernerfolgskontrolle⁶⁵ vorausgesetzt werden. Folglich scheidet die Anwendung dieses Ordnungsprinzips im Sinne der Übertragbarkeit der Ergebnisse aus.

Will man die Cluster vom Konkreten zum Abstrakten anordnen, so ist zunächst eine Definition der Begriffe angebracht (vgl. z.B. [BBAW 2008]). „Konkret“ bedeutet einmal „gegenständlich“, also auf

⁶⁵ Hierbei ist zwischen den Möglichkeiten der Lernplattform – nämlich z.B. wie das WiBA-Net ein Test-System anzubieten – und den speziell für eine Lernerfolgskontrolle erforderlichen Inhalten – nämlich Testfragen und -antworten – zu unterscheiden.

die vorliegende Arbeit bezogen eine Wissensseinheit, die etwas körperlich Greifbares zum Inhalt hat. Eine weitere Bedeutung ist „genau“. Eine konkrete Wissensseinheit kann im Rahmen der vorliegenden Arbeit also auch als einen genauen Bezug auf die ursprünglichen Suchbegriffe aufweisend interpretiert werden. Eine abstrakte Wissensseinheit hat demgegenüber einen „begrifflichen“ bzw. „unanschaulichen“ Inhalt und keinen direkten Bezug zu den ursprünglichen Suchbegriffen.

Bei der Anordnung vom Konkreten zum Abstrakten werden auch die in die Wissensseinheiten eingebundenen multimedialen Elemente berücksichtigt. Diese können anhand ihrer Dateiendung als Bilder, Filme oder Animationen erkannt werden. Die Grundüberlegung ist, dass Bilder und Filme in der Regel konkrete Inhalte zeigen, während Animationen meist erstellt werden, um abstrakte Sachverhalte darzustellen. Als weitere Möglichkeit, konkrete von abstrakten Inhalten zu unterscheiden, wird das Vorkommen der Original-Suchbegriffe auf den Seiten berücksichtigt. Eine Seite, die einen oder mehrere der Suchbegriffe enthält, kann als konkret auf die Suchanfrage bezogen definiert werden.

Um eine komplexe Wissensseinheit zu definieren, wird der Begriff „komplex“ in der Bedeutung von „vielfältig“ bzw. „zusammengesetzt“ verwendet (vgl. z.B. [BBAW 2008]). Ein solcher Charakter einer Wissensseinheit drückt sich zum einen durch die Darstellung differenzierter Sachverhalte aus, aus denen dann eine Folgerung oder Synthese abgeleitet werden kann. Um Wissensseinheiten, die solche differenzierten Sachverhalte aufweisen, zu erkennen, wird die Höhe des Anteils seltener Begriffe in diesen Wissensseinheiten untersucht. Bei einem hohen Anteil an seltenen Begriffen wird davon ausgegangen, dass darunter vermehrt Fachausdrücke sind, die, wenn sie gehäuft auftreten, für eine höhere Differenziertheit und Komplexität der Wissensseinheit sprechen. Auf der Basis der Begriffsdefinition als „vielfältig“ bzw. „zusammengesetzt“ kann auch die Länge der Wissensseinheiten als Kriterium genutzt werden, eine Wissensseinheit als einfach oder komplex einzuordnen: Eine Wissensseinheit wird als umso komplexer eingeordnet, je mehr Begriffe sie enthält.

Die Anordnung vom Allgemeinen zum Speziellen lässt sich zum einen durch Betrachtung der Kookkurenzen der in den Wissensseinheiten enthaltenen Begriffe abschätzen. Es wird davon ausgegangen, dass ein Begriff, der einen speziellen Sachverhalt bezeichnet, im weiteren Sinne also einen Einzel- oder Sonderfall, nur in wenigen lokalen Kontexten und damit lediglich in einem kleinen globalen Kontext auftritt. Solche Begriffe können an der geringeren Anzahl ihrer Kookkurenzen erkannt werden. Zum anderen kann ein Begriff als „speziell“ charakterisiert werden, wenn es sich dabei um ein Determinativkompositum handelt. Dies liegt schon in der Definition des Determinativkompositums begründet. Der Kopf, also der rechte Teil, eines Kompositums wird durch den linken Teil des Kompositums näher bestimmt, also „spezialisiert“. Dieser Umstand lässt sich nutzen, indem alle Begriffe in einer Wissensseinheit, die einen der ursprünglichen Suchbegriffe als Kopf eines

Kompositums enthalten, als Determinativkomposita betrachtet werden⁶⁶. Je höher der Anteil dieser Determinativkomposita, desto spezieller ist wahrscheinlich der Inhalt der Wissensseinheit in Bezug auf die Suchbegriffe.

Die Anordnung von der Ursache zur Wirkung lässt sich auf Basis der zur Verfügung stehenden Daten nicht realisieren und wird deshalb nicht weiter berücksichtigt.

Eine Zusammenstellung aller für die Zuordnung der Wissensseinheiten zu didaktischen Ordnungsprinzipien verwendeten Kriterien findet sich in Tabelle 5.7. Das Auftreten der angeführten Kriterien bei den Wissensseinheiten kann programmtechnisch leicht erfasst werden, die entsprechende Zuordnung ist ebenfalls eindeutig realisierbar. Jedoch liegt ein nicht zu vernachlässigendes Fehlerpotenzial in den Kriterien selbst. Diese können nur die durchschnittliche Zuordnung bei größeren Mengen von Texten abbilden, aber keine definitive Kategorisierung für jeden Einzelfall ermöglichen. Einzelne Texte können durchaus ein Kriterium für die Einordnung in eine bestimmte Kategorie erfüllen, tatsächlich aber der gegenteiligen Kategorie angehören. Ein Beispiel wäre ein längerer Text, der jedoch durchaus „einfach“ sein kann. Deshalb werden für jede Kategorisierung mindestens zwei Kriterien herangezogen, um diesem Effekt entgegen zu wirken. Ein weiteres Problem bei der eindeutigen Zuordnung ergibt sich durch die Tatsache, dass nicht einzelne Wissensseinheiten kategorisiert werden, sondern aus mehreren Wissensseinheiten zusammengesetzte Cluster. Hierbei ist es möglich, dass sich die Eigenschaften einzelner Wissensseinheiten innerhalb eines Clusters gegenseitig aufheben. Da das Clustering nur auf der Basis inhaltlicher Ähnlichkeit durchgeführt wird, ist es sogar wahrscheinlich, dass Wissensseinheiten aus verschiedenen Kategorien darunter sind. Deshalb wird der Lehrpfad nur in der Variante dem Lerner angeboten, für den sich bei der Bestimmung der Reihenfolge der Cluster die breiteste und damit aussagekräftigste Verteilung ergibt.

Tabelle 5.7: Zusammenstellung für die Auswahl eines Ordnungsprinzips zur Verfügung stehender Kriterien

Ordnungsprinzip	Kriterien für...	Kriterien für...
vom Konkreten zum Abstrakten	<u>...abstrakte Wissensseinheiten:</u> x Animationen vorhanden	<u>...konkrete Wissensseinheiten:</u> x Bilder, Filme vorhanden x ursprüngliche Suchbegriffe vorhanden
vom Einfachen zum Komplexen	<u>...einfache Wissensseinheiten:</u> x kurzer Inhalt x geringer Anteil seltener Begriffe	<u>...komplexe Wissensseinheiten:</u> x umfangreicher Inhalt x hoher Anteil seltener Begriffe

⁶⁶ Unter den gefundenen Komposita können sich prinzipiell auch Kopulativkomposita befinden, deren rechter Konstituent dem jeweiligen ursprünglichen Suchbegriff entspricht. Diese stellen jedoch auch einen speziellen Bedeutungszusammenhang des untersuchten Worts dar und können demzufolge bei der Bestimmung der Determinativkomposita ohne Verfälschung des Ergebnisses berücksichtigt werden.

Ordnungsprinzip	Kriterien für...	Kriterien für...
vom Allgemeinen zum Speziellen	<u>...allgemeine Wissensseinheiten:</u> x geringer Anteil an Begriffen mit wenigen Kookkurrenten	<u>...spezielle Wissensseinheiten:</u> x hoher Anteil an Determinativkomposita x hoher Anteil an Begriffen mit wenigen Kookkurrenten

5.4.3.2 Bestimmung der Kennwerte

Zunächst müssen die Kennwerte, die für die Beurteilung der oben beschriebenen Kriterien benötigt werden, bestimmt werden. Zusätzlich zu den Kennwerten der einzelnen zu untersuchenden Wissensseinheiten werden die jeweils mittleren Größen benötigt, anhand derer die Kennwerte als groß oder klein beurteilt werden können. Die mittleren Größen werden einmalig anhand der gesamten Inhaltsbasis bestimmt.

Die grundlegende Größe ist die Länge der Texte. Diese wird sowohl selbst als Parameter verwendet, als auch bei den anderen Kennwerten als Bezugsgröße genutzt. Für die Längenbestimmung werden die um die Füllworte bereinigte Texte herangezogen. Komposita werden nur als ein Wort gezählt, mehrfach auftretende Worte werden auch mehrfach berücksichtigt. Es wird die summierte Länge für jedes Cluster bestimmt und zusätzlich die durchschnittliche Länge einer Wissensseinheit im Cluster. Die Gesamtlänge dient als Bezugsgröße für die übrigen über ein ganzes Cluster ermittelten Kennwerte. Die durchschnittliche Textlänge l_T wird als Kennwert für die Länge der Datei angesetzt, wenn die Gliederung vom Einfachen zum Komplexen untersucht wird.

Der Anteil von Bild-, oder Filmelementen A_{Bild} bzw. Animationen A_{Anim} , der auf die Textlänge bezogen wird, kann einfach aus dem Quelltext der jeweiligen Dateien ermittelt werden. Bei Bilddateien muss dabei beachtet werden, dass für den Inhalt nicht relevante Bilddateien⁶⁷ nicht mit gezählt werden.

Das Vorhandensein der ursprünglichen Suchbegriffe im Text wird durch den Wert $\text{sim}_{\text{ov}}(d_i)$ für jede Wissensseinheit i berücksichtigt. Die Werte werden über den Cluster summiert und gemittelt. Der Vergleichswert bezieht sich für diese Kenngröße nicht auf alle Dateien, sondern entspricht dem der Relevanzberechnung zugrunde gelegten Mittelwert der gefundenen Dateien (vgl. Abbildung 5.4).

Zur Bestimmung des Anteils der seltenen Begriffe A_{selten} werden alle Worte in den Wissensseinheiten gezählt, die eine Häufigkeit von 1 oder 2 aufweisen⁶⁸. Je höher deren Anzahl, desto höher ist der Anteil seltener Begriffe an einer Wissensseinheit.

⁶⁷ z.B. Logos oder auch Symbole, die als Kennzeichen eines vergrößerbaren multimedialen Elements verwendet werden.

⁶⁸ Die Ermittlung einer mittleren Häufigkeit aller Begriffe einer Wissensseinheit ist nicht ausreichend, da hierbei das Auftreten seltener Begriffe durch das gleichzeitige Auftreten häufiger Begriffe wieder ausgeglichen werden kann.

Der Anteil der Determinativkomposita am Gesamttext A_{Det} ist die einzige der Kenngrößen, deren Vergleichswert anhand des zu untersuchenden Lehrpfads bestimmt werden muss. Der Grund hierfür ist, dass nur Determinativkomposita der ursprünglichen Suchbegriffe berücksichtigt werden, die nicht vorab bekannt sind, sodass der Wert nicht zentral vorgegeben werden kann. Eine zur Lehrpfaderstellung parallele Ermittlung über alle Wissensseinheiten scheidet aus Effizienzgründen aus. Der Vergleichswert wird aus den für alle Cluster des Lehrpfads gefundenen Determinativkomposita ermittelt.

Um den Anteil der Begriffe mit wenigen Kookkurrenten A_{spez} zu ermitteln, werden für jedes Cluster alle Begriffe gezählt, die weniger als 100 Kookkurrenten aufweisen. Die Ermittlung einer mittleren Kookkurrentenzahl pro Cluster ist nicht sinnvoll, da einige, insbesondere häufig auftretende Begriffe, eine große Anzahl an Kookkurrenten aufweisen, die das durchschnittliche Ergebnis verfälschen würden.

5.4.3.3 Bestimmung der Varianten

Die Zuordnung zu den einzelnen Kategorien wird mit drei Fuzzy-Berechnungen durchgeführt. Dies bietet sich aufgrund der nicht gegebenen mathematischen Zusammenhänge zwischen den Ein- und Ausgangsgrößen an. Die Fuzzyifizierung erfolgt dabei für die einzelnen scharfen Eingangsgrößen nach einem für alle Gliederungsvarianten festen Schema. Die Inferenzregeln für die einzelnen Gliederungen folgen aus den beschriebenen Eigenschaften eines Clusters in Abhängigkeit von den entsprechenden Eingangsgrößen. Die Defuzzyifizierung nach der Methode der Teilschwerpunkte erfolgt abschließend wieder mit einem gemeinsamen Schema. Im Folgenden wird der Berechnungsablauf erläutert.

Bei der Fuzzyifizierung werden dreiecks- bzw. trapezförmige Zugehörigkeitsfunktionen angewendet. Ihr prinzipieller Verlauf ist Abbildung 5.9 zu entnehmen.

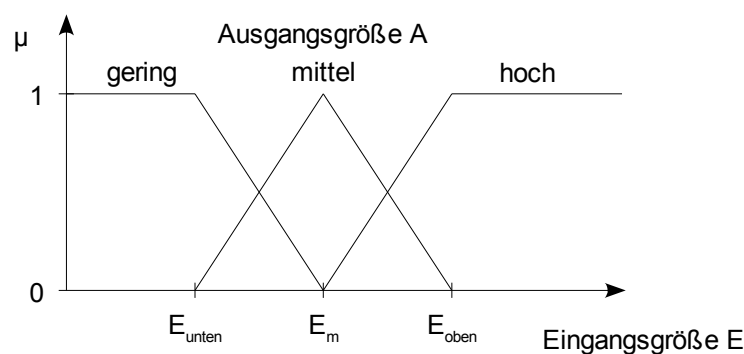


Abbildung 5.9: Prinzipieller Verlauf der Zugehörigkeitsfunktionen der Eingangsgrößen

Insgesamt werden sieben scharfe Eingangsgrößen fuzzyfiziert. Für jede Eingangsgröße werden die Kennwerte der Zugehörigkeitsfunktionen benötigt. Dies ist zum einen der Wert für das Maximum der Zugehörigkeitsfunktion für die mittlere Ausgangsgröße (E_m), der Wert, bei dem die geringe Ausgangsgröße in den linearen Bereich übergeht (E_{unten}) und der Wert, bei dem die hohe Eingangsgröße den konstanten Bereich erreicht (E_{oben}). Für die Eingangsgrößen wird beim Anteil der Bilder A_{Bild} , dem der Animationen $A_{Anim.}$, der durchschnittlichen Textlänge I_T , dem Anteil an seltenen Begriffen A_{selten} und dem Anteil an Begriffen mit wenigen Kookkurrenten $A_{spez.}$ jeweils der Mittelwert für alle Dateien der Inhaltsbasis bestimmt und als E_m übergeben. Für E_{oben} bzw. E_{unten} werden die Werte gewählt, oberhalb bzw. unterhalb derer jeweils 2 % aller Werte liegen. Für den Anteil der Determinativkomposita $A_{Det.}$ wird wie oben beschrieben der Mittelwert aus den Clustern eines Lehrpfads als E_m übergeben, als E_{oben} wird der maximale Anteilwert, als E_{unten} der minimale übergeben. Für die Ähnlichkeit zu den ursprünglichen Suchbegriffen wird als E_{unten} der Mittelwert der schon für die Relevanzberechnung verwendeten Werte $sim_{OV}(d_i)$ verwendet, als E_{oben} der Wert $W_{99,5;OV}$ und als E_m der Mittelwert aus diesen Werten. Aus diesen Eingangswerten können mit Hilfe der für die jeweilige Wissensseinheit d_i vorhandenen Werte die Zugehörigkeitsgrade zu den unscharfen Mengen bestimmt werden. Eine Übersicht findet sich in Tabelle 5.8.

Tabelle 5.8: Zugehörigkeitsgrade zu den unscharfen Mengen und Parameter für die Zugehörigkeitsfunktionen für die Clustersortierung

Merkmal	Ähnlichkeit / Anteil			Parameter der Zugehörigkeitsfunktion		
	gering	mittel	hoch	E_{unten}	E_m	E_{oben}
I_T	$\mu_{I,gering}(d_i)$	$\mu_{I,mittel}(d_i)$	$\mu_{I,hoch}(d_i)$	25,0	73,831	129,0
A_{Bild}	$\mu_{Bild,gering}(d_i)$	$\mu_{Bild,mittel}(d_i)$	$\mu_{Bild,hoch}(d_i)$	0,0	1,183	3,0
$A_{Anim.}$	$\mu_{Anim.,gering}(d_i)$	$\mu_{Anim.,mittel}(d_i)$	$\mu_{Anim.,hoch}(d_i)$	0,0	0,102	1,0
A_{selten}	$\mu_{selten,gering}(d_i)$	$\mu_{selten,mittel}(d_i)$	$\mu_{selten,hoch}(d_i)$	0,0	0,096	0,264
$A_{spez.}$	$\mu_{spez.,gering}(d_i)$	$\mu_{spez.,mittel}(d_i)$	$\mu_{spez.,hoch}(d_i)$	0,037	0,219	0,5
$A_{Det.}$	$\mu_{Det.,gering}(d_i)$	$\mu_{Det.,mittel}(d_i)$	$\mu_{Det.,hoch}(d_i)$	abhängig von der Suchanfrage		
$sim_{OV}(d_i)$	$\mu_{OV,gering}(d_i)$	$\mu_{OV,mittel}(d_i)$	$\mu_{OV,hoch}(d_i)$	abhängig von der Suchanfrage		

Die Inferenzregeln für die einzelnen Fuzzy-Berechnung folgen wegen der begrenzten Anzahl an Eingangswerten anschaulich aus den oben beschriebenen Überlegungen zur Charakterisierung der Eigenschaften von Seiten durch textbasierte Kennwerte. Die entsprechenden Verarbeitungsregeln können den Tabellen 5.9 bis 5.11 entnommen werden. Die Eingangswerte sind für geringe (g), mittlere (m) und hohe (h) Zugehörigkeiten als Spalten- bzw. Zeilenköpfe aufgetragen.

Tabelle 5.9: Inferenzregeln für die Sortierung eines Lehrpfads vom Konkreten (-) zum Abstrakten (+)

Zeile	sim _{ov}	A _{Bild} \ A _{Anim.}	g	m	h
1	h	h	+	+	0
2	h	m	+	0	0
3	h	g	+	0	0
4	m	h	+	0	0
5	m	m	0	0	-
6	m	g	0	0	-
7	g	h	0	0	-
8	g	m	0	0	-
9	g	g	0	-	-

Tabelle 5.10: Inferenzregeln für die Sortierung eines Lehrpfads vom Einfachen (-) zum Komplexen (+)

Zeile	A _{selten} \ I _{r.}	g	m	h
1	h	+	+	+
2	m	0	0	0
3	g	-	-	0

Tabelle 5.11: Inferenzregeln für die Sortierung eines Lehrpfads vom Allgemeinen (-) zum Speziellen (+)

Zeile	A _{Det.} \ A _{spez.}	g	m	h
1	h	0	+	+
2	m	-	0	+
3	g	-	0	+

Als Werte finden sich die Schlussfolgerungen, die sich aus dem Eintreten der Prämissen ergeben. Hierbei sind jeweils ein hoher Wert (+), ein mittlerer (0) und ein geringer (-) möglich. Ein hoher Wert steht dabei für einen ausgeprägten abstrakten bzw. komplexen bzw. speziellen Charakter einer Wissenseinheit. Ein niedriger Wert steht demgegenüber für einen konkreten bzw. einfachen bzw. allgemeinen Charakter. Die Berechnung der unscharfen Ausgangswerte läuft analog zum Beispiel in Abschnitt 5.3.3 ab.

Aus den ermittelten unscharfen Ausgangsgrößen werden durch die Defuzzifizierung wieder scharfe Werte ermittelt. Hierbei wird die Methode der Teilschwerpunkte angewendet. Die nach dem jeweiligen Ordnungsprinzip zu sortierenden Cluster werden in der Reihenfolge der aufsteigenden scharfen Ausgangswerte hintereinander platziert.

Abschließend ist die Gliederung mit der höchsten Aussagekraft auszuwählen. Als Kriterium für die Aussagekraft wird, wie oben erläutert, die Breite der Verteilung gewählt. Diese wird durch die Standardabweichung charakterisiert. Den Clustern wird für jede Gliederungsvariante als Ergebnis der jeweiligen Fuzzy-Berechnung ein Wert einer Skala zwischen 0 und 100 zugewiesen, der eine Einordnung in der entsprechenden Kategorie ermöglicht. Bei einem aus n Clustern bestehenden Lehrpfad ergeben sich also je Gliederungsvariante n Skalenwerte. Aus diesen werden der Mittelwert und die Standardabweichung bestimmt. Die Gliederung mit der größten Standardabweichung wird als aussagekräftigste interpretiert. Je nach Nutzer wird das Ergebnis unterschiedlich angezeigt: Lehrende erhalten Zugriff auf die drei Versionen mit Informationen zur jeweiligen Standardabweichung. Lernenden wird nur die ausgewählte Version angezeigt. Ihnen wird mitgeteilt, welchem Prinzip die Gliederungsvariante unterliegt, sodass die Bearbeitung in diesem Kontext stattfinden kann.

5.4.4 Hierarchische Gliederung

Die Cluster werden im letzten Schritt intern hierarchisch gegliedert, bleiben aber hinsichtlich der enthaltenen Seiten unverändert. Jeder Cluster wird in eine Hauptseite und Nebenseiten unterteilt, die Struktur des Lehrpfads beschränkt sich also auf zwei Ebenen. Um eine weitere Gliederung zu erreichen, müsste der Prozess des Clusters, Sortierens und Gliederns für jede Ebene wiederholt werden. Hierauf soll im Rahmen der vorliegenden Arbeit verzichtet werden, da die anzuwendenden Grundprinzipien die gleichen sind. Außerdem wird bei einem tiefer strukturierten Lehrpfad die Auswertung der Ergebnisse unanschaulicher und die den erzeugten Lehrpfaden zugrunde liegenden Prinzipien sind schwerer nachvollziehbar.

Im Folgenden werden Kriterien diskutiert, die aus einer Inhaltsseite automatisch gewonnen und für die hierarchische Zuordnung dieser Seite bzw. Wissensseinheit in einem Cluster genutzt werden können. Dabei wird erläutert, wie die Seite auf der Basis des jeweiligen Kriteriums beurteilt werden kann. Die Kriterien lassen sich gliedern in solche, die nur auf dem eigentlichen Text der Wissensseinheit basieren und in solche, die auch den Kontext zugrunde legen. Für die Zuordnung einer Wissensseinheit i zu den Hierarchieebenen „Hauptseite“ oder „Nebenseite“ wird ein gewichteter Hierarchiewert H_i berechnet, der auf Zahlenwerten basiert, die sich aus den folgenden Kriterien ableiten lassen. Die Wissensseinheit, die in einem Cluster den höchsten Wert H_i erreicht, wird zur Hauptseite dieses Clusters, die übrigen Seiten bilden die Nebenseiten.

Textbasierte Kriterien

1. Die Original-Suchterme (d.h. die vom Nutzer eingegebenen) kommen im Text der Seite einmal oder mehrmals vor. Generell spricht ein Vorkommen der Original-Terme für einen Status als Hauptseite, also einen hohen H_i -Wert. Ein Vorkommen von ausschließlich aus der Erweiterung stammenden Suchtermen spricht eher für eine Nebenseite, also einen niedrigen H_i -Wert.
2. Die Original-Suchterme kommen in der Überschrift vor. Dies ist ein Kriterium für eine Einordnung als Hauptseite. Wenn eine Seite einen der Suchbegriffe als alleinigen Term in der Überschrift enthält, führt dies zu einem hohen H_i -Wert. Dies ist ebenso der Fall, wenn eine Seite mehrere Suchbegriffe im Titel führt. Eine aus mehreren Worten bestehende Überschrift mit nur einem Suchbegriff wird in der Hierarchie tiefer eingeordnet.
3. Die Seite hat eine hohe Ähnlichkeit mit den Original-Suchvektoren. Diese Ähnlichkeit entspricht dem sim_{OV} -Wert. Je höher diese Ähnlichkeit ist, desto höher ist der H_i -Wert. Eine Seite mit einer hohen Ähnlichkeit stellt eine direkte Verbindung zum Thema der Suche dar und kommt deshalb als Hauptseite in Frage. Seiten niedriger Ähnlichkeit eignen sich eher als Nebenseiten.

Kontextbasierte Kriterien

4. Die Seite ist den übrigen Seiten im Cluster möglichst ähnlich. Für den Dokumentvektor jeder Seite im Cluster wird der Abstand („euklidische Distanz“) zu den anderen Dokumentvektoren bestimmt. Die erhaltenen Werte werden gemittelt. Je kleiner die mittlere Distanz zu allen anderen Seiten ist, desto größer wird der H_i -Wert. Eine Seite, die im Mittel sehr weit von den anderen entfernt ist, kann ggf. als thematischer Ausreißer betrachtet werden. Ihr Beitrag zum H_i -Wert ist deshalb gering.
5. Es handelt sich um eine Seite mit einem Inhalt zu einem speziellen Thema. Hauptseiten sind den Nebenseiten übergeordnet, auf ihnen wird in der Regel ein allgemeines Thema behandelt, das durch die Nebenseiten vertieft wird. Auf den Nebenseiten sind deshalb eher spezielle Themen zu finden. Eine Nebenseite ist deshalb, analog zur Sortierung des Lehrpfads⁶⁹, durch einen hohen Anteil von Begriffen mit wenigen Kookkurrenten charakterisiert. Ein hoher Anteil führt also zu einem niedrigen H_i -Wert.

Für die genannten Kriterien werden Zahlenwerte bestimmt. Diese lassen sich für die Kriterien 1 bis 3 für eine Wissensseinheit bzw. ein Dokument d_i als $A_{FT}(d_i)$, $A_{UE}(d_i)$ und $\text{sim}_{OV}(d_i)$ aus der Relevanz-

⁶⁹ vgl. Abschnitt 5.4.3

berechnung übernehmen. Diese Werte haben die Eigenschaft, dass bei betragsmäßig größeren Werten auch H_i größer werden muss. Ferner nehmen sie nur Zahlenwerte zwischen 0 und 1 an.

Für das Kriterium 4 wird der mittlere Abstand $a_{m,i}$ einer Wissensseinheit i zu den übrigen Wissensseinheiten im Cluster als Zahlenwert herangezogen. Kriterium 5 wird durch den Anteil von Begriffen $A_{<100,i}$ mit weniger als 100 Konkurrenten am Gesamttextumfang der Wissensseinheit i repräsentiert. Den letzten beiden Kriterien ist gemein, dass der zu bestimmende H_i -Wert größer werden muss, wenn der jeweilige Zahlenwert kleiner wird. Des Weiteren ist ihr Wertebereich nicht auf das Intervall $[0,1]$ beschränkt, sondern sie können quasi beliebige positive Werte annehmen. Aus diesem Grund müssen die Zahlenwerte für die Kriterien 4 und 5 normiert werden.

Damit der Wert H_i ansteigt wenn die Zahlenwerte der Kriterien 4 und 5 ansteigen, werden stattdessen die Kehrwerte $1/a_{m,i}$ und $1/A_{<100,i}$ betrachtet. Diese werden auf ihren jeweils größten Wert, also den Kehrwert des minimalen im Cluster auftretenden Zahlenwerts, bezogen, sodass gilt:

$$\frac{1}{a_{m,i}} = \frac{a_{m,min}}{a_{m,i}} \quad (5.12)$$

bzw.

$$\frac{1}{A_{<100,i}} = \frac{A_{<100,min}}{A_{<100,i}} \quad (5.13)$$

Alle ermittelten Werte nehmen also Werte zwischen 0 und 1 an. Ein Anstieg der Werte führt auch zu einem Anstieg von H_i . Die kontextbasierten Werte sind für die Einstufung einer Seite als Haupt- oder Nebenseite am wichtigsten und werden deshalb mit dem Faktor 1 gewichtet. Die drei textbasierten Kriterien, die lediglich das Vorkommen der ursprünglichen Suchbegriffe berücksichtigen, werden mit dem Faktor 1/3 gewichtet. Für H_i gilt folglich:

$$H_i = \frac{1}{a_{m,i}} + \frac{1}{A_{<100,i}} + \frac{A_{FT}(D_i)}{3} + \frac{A_{UE}(D_i)}{3} + \frac{sim_{OV}(D_i)}{3} \quad (5.14)$$

Der Wert H_i kann also Werte zwischen 0 und 3 annehmen. Die Wissensseinheit i , deren H_i -Wert am größten ist, wird Hauptseite ihres Clusters.

Mit der Festlegung der hierarchischen Gliederung der Cluster ist die automatische Erstellung des Lehrpfads abgeschlossen. Der fertige Lehrpfad wird gespeichert und ist jetzt zur Bearbeitung durch Lernende oder Lehrende bereit. Er bildet die Ergebnismenge E_2 .

6 Ergebnisse der Lehrpfaderstellung

6.1 Lehrpfade als Ergebnis von Suchanfragen

Zur Veranschaulichung der bislang dargestellten theoretischen Zusammenhänge soll an dieser Stelle zunächst ein exemplarischer Lehrpfad, der mit dem implementierten Programmsystem erstellt wurde, vorgestellt werden. In Tabelle 6.1 sind die enthaltenen Wissensseinheiten und die Struktur des Lehrpfads zu der Suchanfrage „Gleichgewichtsfeuchte – Holz – Quellen“ dargestellt. Die einzelnen Lehrpfadseiten finden sich in Anhang I. Gegenüber dem in Abschnitt 3.4.2 dargestellten WiBA-Net-Lehrpfad „Konsistenzbestimmung“ treten beim automatisch generierten Lehrpfad außerdem weiterführende Seiten auf, die in der rechten Spalte der Tabelle dargestellt sind.

Tabelle 6.1: Aufbau des Lehrpfads zu den Suchbegriffen „Gleichgewichtsfeuchte – Holz – Quellen“

Nr.	Hauptseiten	Nebenseiten	weiterführende Seiten
1	Einfluss der Temperatur auf das Verformungsmodul von Holz		
2		Einfluss der Holzfeuchte auf das Verformungsverhalten	<i>Verformungsverhalten von Holz unter Kurzzeitbelastung</i>
3	Gleichgewichtsfeuchte von Holz		<i>Temperaturabhängigkeit der Holzfestigkeit</i> <i>Abhängigkeit der Festigkeit von der Rohdichte</i>
4		Schwinden und Quellen von Holz	
5	Trocknen von Holz		
6		Sorptionsisotherme und Mechanismen der Feuchtaufnahme	
7	Charakteristische Eigenschaften – Schwinden und Quellen		
8		Quell- und Schwindmaße von Holz	
9		Relevante Eigenschaften der für Bauzwecke eingesetzten Holzarten	
10		Schwindverformungen	

Man erkennt, dass das Thema, das sich um die Suchbegriffe aufspannt, von dem erstellten Lehrpfad recht gut abgedeckt wird. Bei unterschiedlichen automatisch erstellten Lehrpfaden kann die inhaltliche und didaktische Qualität jedoch variieren. In den folgenden Abschnitten werden deshalb der hier vorgestellte und zwei weitere Lehrpfade detailliert untersucht. Dabei werden insbesondere die Zwischenergebnisse und der Ablauf der Berechnung analysiert, unter dem Gesichtspunkt, wie einzelne Seiten in einen Lehrpfad und in diesem Lehrpfad auf eine bestimmte Position gelangen

und welche Einflüsse auf die inhaltliche Qualität eines automatisch erstellten Lehrpfads existieren. Für die Lehrpfade 2 und 3 werden die Inhalte selbst nicht als Screenshots abgebildet, die Darstellung beschränkt sich hier auf die Überschriften der einzelnen Seiten.

6.2 Analyse automatisch erstellter Lehrpfade

Im folgenden sollen anhand von drei automatisch erstellten Lehrpfaden zu verschiedenen Suchanfragen und daraus abgeleiteten Alternativen die Erstellung von Lehrpfaden sowie dabei auftretende Besonderheiten erläutert werden. Es wird Wert auf den Vergleich der Ergebnisse bei einer Variation der Eingangswerte gelegt. Für den ersten der erstellten Lehrpfade erfolgt eine detaillierte Analyse aller Schritte der automatischen Lehrpfaderstellung und der jeweils berechneten Zwischenergebnisse. Beim zweiten und dritten Lehrpfad werden darauf aufbauend spezielle Aspekte betrachtet, welche die Nutzungsaspekte des Systems betreffen.

Besonders interessieren dabei Wissensseinheiten, die, obwohl sie für den Lehrpfad relevant wären, nicht in diesen aufgenommen wurden, und solche, die Bestandteil des Lehrpfads sind, obwohl sie für diesen irrelevant sind. In diesen Fällen wird nachvollzogen, durch welchen Berechnungsweg sie jeweils außerhalb bzw. innerhalb des Lehrpfads platziert wurden.

Betrachtet werden Lehrpfade zu den folgenden Suchbegriffen:

- Lehrpfad 1: Gleichgewichtsfeuchte – Holz – Quellen
- Lehrpfad 2: Polykondensation – Polymerisation
- Lehrpfad 3: Hydratation – Zement

In den folgenden Abschnitten findet sich am Anfang jeweils eine Übersicht über den generierten Lehrpfad in Tabellenform. Darin sind die Nebenseiten der Cluster eingerückt sowie die weiterführenden Seiten eingerückt und kursiv dargestellt. Das ausgewählte Ordnungsprinzip ist angegeben. Es wurden in jedem der Lehrpfade nach der Relevanzberechnung entsprechend dem in Abschnitt 5.3.5 beschriebenen Verfahren Seiten ergänzt und entfernt, weshalb berechneter und tatsächlicher Umfang des Lehrpfads differieren. Dargestellt ist jeweils der endgültige Zustand des Lehrpfads unter Berücksichtigung der ergänzten und entfernten Seiten.

6.3 Lehrpfad 1

Die Suchanfrage „Gleichgewichtsfeuchte – Holz – Quellen“ beschäftigt sich mit der Thematik der Formänderung von Holz in Abhängigkeit von der Holz- bzw. Umgebungsfeuchte. Dabei wurde als Lernziel eines fiktiven Lernalters unterstellt, den Zusammenhang zwischen der Gleichgewichtsfeuch-

te und dem Quellen von Holz zu verstehen. Der erstellte Lehrpfad besteht aus 10 Seiten, die zu vier Clustern zusammengefasst sind. Ursprünglich wurden mit der Relevanzberechnung 11 Seiten ermittelt. Davon wurden zwei Seiten verworfen, da sie eine zu geringe Ähnlichkeit mit den Unterthemen aufwiesen⁷⁰. Eine Seite wurde wegen einer entsprechend hohen Ähnlichkeit ergänzt⁷¹. Der Lehrpfad entspricht dem in Abschnitt 6.1 vorgestellten Lehrpfad. Seine Struktur kann Tabelle 6.2 entnommen werden.

Tabelle 6.2: Zusammenstellung der untersuchten Lehrpfade

Suchbegriffe / Kenngrößen	Struktur des erzeugten Lehrpfads
<ul style="list-style-type: none"> ● Gleichgewichtsfeuchte ● Holz ● Quellen <p>Sollwert L: 11 Seiten</p> <p>tatsächlicher Umfang: 10 Seiten</p> <p>Ordnungsprinzip: vom Allgemeinen zum Speziellen</p>	<ul style="list-style-type: none"> ► Einfluss der Temperatur auf das Verformungsmodul von Holz <ul style="list-style-type: none"> - Einfluss der Holzfeuchte auf das Verformungsverhalten <ul style="list-style-type: none"> - <i>Verformungsverhalten von Holz unter Kurzzeitbelastung</i> ► Gleichgewichtsfeuchte von Holz <ul style="list-style-type: none"> - <i>Temperaturabhängigkeit der Holzfestigkeit</i> - <i>Abhängigkeit der Festigkeit von der Rohdichte</i> - Schwinden und Quellen von Holz ► Trocknen von Holz <ul style="list-style-type: none"> - Sorptionsisotherme und Mechanismen der Feuchtaufnahme ► Charakteristische Eigenschaften – Schwinden und Quellen <ul style="list-style-type: none"> - Quell- und Schwindmaße von Holz - Relevante Eigenschaften der für Bauzwecke eingesetzten Holzarten - Schwindverformungen
Kennzeichnung der Strukturierung: ► Hauptseiten, - Nebenseiten, - <i>weiterführende Seiten</i>	

Der zum Auffinden der Seiten benötigte erweiterte Suchvektor besteht aus den in der Tabelle 6.3 angegebenen Begriffen.

Tabelle 6.3: Suchtermerweiterungen für die Suchbegriffe „Gleichgewichtsfeuchte“ – „Holz“ – „Quellen“

Grundlage der Erweiterung	Suchbegriffe
Original-Suchbegriffe	gleichgewichtsfeucht, holz, quell
paarweise gemeinsame ähnliche Begriffe	angegeben, baustofftemperatur, eingegeben, fasersaettigung, feucht, feuchtegehalt, feuchtegehaltsaender, formhaltigkeit, holzfeucht, ugl, verformungsaenderungen
gemeinsame Kookkurrenten	feucht, holzbau, schwind, ugl
höchstsignifikante Kookkurrenten der Original-Suchbegriffe	aender, anzustreb, baupraktisch, baustoff, eigenschaf ⁷² , einbausituation, feucht, feuchtegehalt, feuchtegehaltsaender, geschlagen, gewicht,

70 „Stromquellen“ und „Erklärung der Kurzzeichen und Fachbegriffe (Holz)“

71 „Einfluss der Temperatur auf das Verformungsmodul von Holz“

72 Diese kürzeste Wortform hätte nach den in Abschnitt 4.2.3 erläuterten Regeln eigentlich nicht für den Begriff „Eigenschaft“ gefunden werden dürfen. Jedoch sind im Textkorpus des WiBA-Net gelegentlich Rechtschreibfehler zu finden, die Form „eigenschaf“ ergibt sich nach den o.a. Regeln aus dem falsch geschriebenen Wort „Eigenschaften“.

	gleich, gleichgewicht, holz, holzbau, holzquerschnitt, holzschutz, informationsdienst, klima, schweisstromquelle, schwind, schwindverhalten, sorptionsisotherme, strom, stromquell, trocken, ugl, verarbeitet, wechselstrom, werkstoff
--	--

Auf der Basis dieser Suchbegriffe wurden vom System 1185 Wissensseinheiten gefunden. Für die Relevanzberechnung auf Fuzzy-Basis wurden für alle Wissensseinheiten die Ähnlichkeitswerte sim_{SV} , sim_{FV} und sim_{OV} bestimmt. Für die Fuzzifizierung ergaben sich folgende Mittelwerte und $W_{99,5}$ -Werte:

$$\text{sim}_{SV,m} = 0,052; \quad W_{99,5;SV} = 0,296$$

$$\text{sim}_{FV,m} = 0,172; \quad W_{99,5;FV} = 0,640$$

$$\text{sim}_{OV,m} = 0,018; \quad W_{99,5;OV} = 0,242$$

und damit die in den Abbildungen 6.1 bis 6.3 dargestellten Zugehörigkeitsfunktionen.

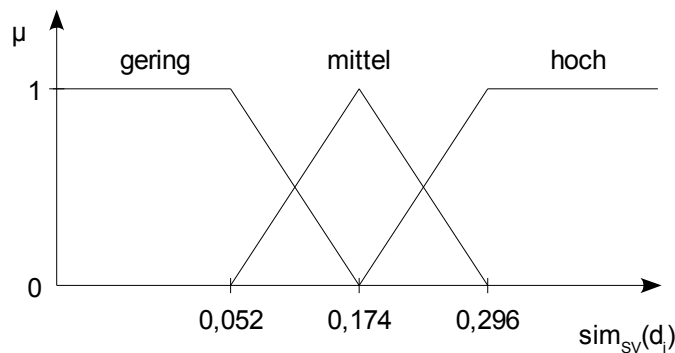


Abbildung 6.1: Zugehörigkeitsfunktionen für die Ähnlichkeit des erweiterten Suchvektors $\text{sim}_{SV}(d_i)$ mit den Wissensseinheiten

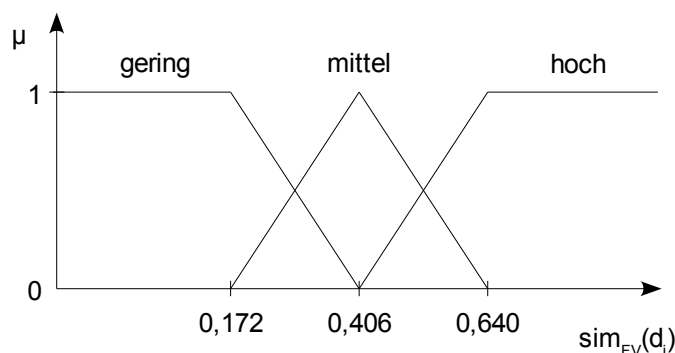


Abbildung 6.2: Zugehörigkeitsfunktionen für die Ähnlichkeit des erweiterten Vektors der Suchbegriffe, die in der Wissensseinheit vorkommen $\text{sim}_{FV}(d_i)$ mit den Wissensseinheiten

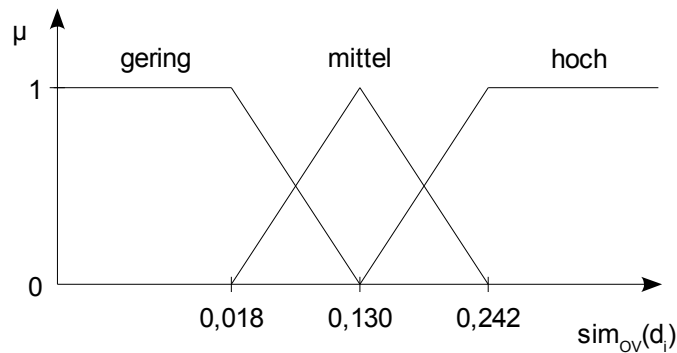


Abbildung 6.3: Zugehörigkeitsfunktionen für die Ähnlichkeit der ursprünglichen Suchbegriffe $sim_{OV}(d_i)$ mit den Wissensseinheiten

Die Werte und Zugehörigkeitsfunktionen für die Anteilswerte A_{FT} und A_{UE} sind unabhängig von der Suchanfrage und wurden in Abschnitt 5.3.2 definiert (vgl. Abbildung 5.5).

Für alle Wissensseinheiten sind die berechneten scharfen Eingangswerte, die sich daraus ergebenden unscharfen Eingangswerte sowie der scharfe Ausgangswert, die Relevanz R , in Tabelle 6.4 angegeben. Die Spalte $sim_{UT,n}(d_i)$ gibt dabei die Ähnlichkeiten der einzelnen Wissensseinheiten zu den Suchwort-Kombinationen nach Formel (5.9) zur Berücksichtigung von Unterthemen an. Hierbei steht „g“ für „Gleichgewichtsfeuchte“, „h“ für Holz und „q“ für „Quellen“.

Tabelle 6.4: Kennwerte der Relevanzberechnung der Wissensseinheiten

Rang / Relevanz R	Titel	$sim_{SV}(d_i) /$ μ_{SV}	$sim_{FV}(d_i) /$ μ_{FV}	$sim_{OV}(d_i) /$ μ_{OV}	A_{FT} μ_{FT}	A_{UE} μ_{UE}	$sim_{UT,n}(d_i)$
1. 89,17	Schwinden und Quellen von Holz	0,393 $\mu_{SV,gering} = 0,0$ $\mu_{SV,mittel} = 0,0$ $\mu_{SV,hoch} = 1,0$	0,625 $\mu_{FV,gering} = 0,0$ $\mu_{FV,mittel} = 0,07$ $\mu_{FV,hoch} = 0,93$	0,289 $\mu_{OV,gering} = 0,0$ $\mu_{OV,mittel} = 0,0$ $\mu_{OV,hoch} = 1,0$	100 % $\mu_{FT,gering} = 0,0$ $\mu_{FT,mittel} = 0,0$ $\mu_{FT,hoch} = 1,0$	66,67 % $\mu_{UE,gering} = 0,0$ $\mu_{UE,mittel} = 0,0$ $\mu_{UE,hoch} = 1,0$	g – h: 0,349 g – q: 0,313 h – q: 0,455
2. 89,17	Gleichge- wichtsfeuchte von Holz	0,294 $\mu_{SV,gering} = 0,0$ $\mu_{SV,mittel} = 0,02$ $\mu_{SV,hoch} = 0,98$	0,476 $\mu_{FV,gering} = 0,0$ $\mu_{FV,mittel} = 0,69$ $\mu_{FV,hoch} = 0,31$	0,4 $\mu_{OV,gering} = 0,0$ $\mu_{OV,mittel} = 0,0$ $\mu_{OV,hoch} = 1,0$	100 % $\mu_{FT,gering} = 0,0$ $\mu_{FT,mittel} = 0,0$ $\mu_{FT,hoch} = 1,0$	66,67 % $\mu_{UE,gering} = 0,0$ $\mu_{UE,mittel} = 0,0$ $\mu_{UE,hoch} = 1,0$	g – h: 0,6 g – q: 0,594 h – q: 0,447
3. 89,17	Quell- und Schwindmaße von Holz	0,347 $\mu_{SV,gering} = 0,0$ $\mu_{SV,mittel} = 0,0$ $\mu_{SV,hoch} = 1,0$	0,619 $\mu_{FV,gering} = 0,0$ $\mu_{FV,mittel} = 0,08$ $\mu_{FV,hoch} = 0,92$	0,242 $\mu_{OV,gering} = 0,0$ $\mu_{OV,mittel} = 0,01$ $\mu_{OV,hoch} = 0,99$	66,67 % $\mu_{FT,gering} = 0,0$ $\mu_{FT,mittel} = 0,0$ $\mu_{FT,hoch} = 1,0$	66,67 % $\mu_{UE,gering} = 0,0$ $\mu_{UE,mittel} = 0,0$ $\mu_{UE,hoch} = 1,0$	g – h: 0,189 g – q: 0,333 h – q: 0,5
4. 82,25	Sorptionsiso- therme und Mechanismen der Feuchte- aufnahme	0,313 $\mu_{SV,gering} = 0,0$ $\mu_{SV,mittel} = 0,0$ $\mu_{SV,hoch} = 1,0$	0,555 $\mu_{FV,gering} = 0,0$ $\mu_{FV,mittel} = 0,37$ $\mu_{FV,hoch} = 0,63$	0,302 $\mu_{OV,gering} = 0,0$ $\mu_{OV,mittel} = 0,0$ $\mu_{OV,hoch} = 1,0$	66,67 % $\mu_{FT,gering} = 0,0$ $\mu_{FT,mittel} = 0,0$ $\mu_{FT,hoch} = 1,0$	0 % $\mu_{UE,gering} = 1,0$ $\mu_{UE,mittel} = 0,0$ $\mu_{UE,hoch} = 0,0$	g – h: 0,543 g – q: 0,357 h – q: 0,354
5. 72,58	Einfluss der Holzfeuchte auf das Verfor- mungsverhalten	0,286 $\mu_{SV,gering} = 0,0$ $\mu_{SV,mittel} = 0,09$ $\mu_{SV,hoch} = 0,91$	0,6 $\mu_{FV,gering} = 0,0$ $\mu_{FV,mittel} = 0,18$ $\mu_{FV,hoch} = 0,82$	0,234 $\mu_{OV,gering} = 0,0$ $\mu_{OV,mittel} = 0,08$ $\mu_{OV,hoch} = 0,92$	33,33 % $\mu_{FT,gering} = 0,33$ $\mu_{FT,mittel} = 0,67$ $\mu_{FT,hoch} = 0,0$	33,33 % $\mu_{UE,gering} = 0,33$ $\mu_{UE,mittel} = 0,67$ $\mu_{UE,hoch} = 0,0$	g – h: 0,397 g – q: 0,179 h – q: 0,183

Rang / Relevanz R	Titel	$\text{sim}_{SV}(d_i) / \mu_{SV}$	$\text{sim}_{FV}(d_i) / \mu_{FV}$	$\text{sim}_{OV}(d_i) / \mu_{OV}$	A_{FT} / μ_{FT}	A_{UE} / μ_{UE}	$\text{sim}_{UT,n}(d_i)$
6. 71,46	<i>Charakteristische Eigenschaften – Schwinden und Quellen</i>	0,237 $\mu_{SV,gering} = 0,0$ $\mu_{SV,mittel} = 0,49$ $\mu_{SV,hoch} = 0,51$	0,565 $\mu_{FV,gering} = 0,0$ $\mu_{FV,mittel} = 0,32$ $\mu_{FV,hoch} = 0,68$	0,191 $\mu_{OV,gering} = 0,0$ $\mu_{OV,mittel} = 0,46$ $\mu_{OV,hoch} = 0,54$	66,67 % $\mu_{FT,gering} = 0,0$ $\mu_{FT,mittel} = 0,0$ $\mu_{FT,hoch} = 1,0$	33,33 % $\mu_{UE,gering} = 0,33$ $\mu_{UE,mittel} = 0,67$ $\mu_{UE,hoch} = 0,67$	$g - h: 0,2$ $g - q: 0,247$ $h - q: 0,404$
7.* 67,03	<i>Stromquellen</i>	0,205 $\mu_{SV,gering} = 0,0$ $\mu_{SV,mittel} = 0,75$ $\mu_{SV,hoch} = 0,25$	0,486 $\mu_{FV,gering} = 0,0$ $\mu_{FV,mittel} = 0,65$ $\mu_{FV,hoch} = 0,35$	0,280 $\mu_{OV,gering} = 0,0$ $\mu_{OV,mittel} = 0,0$ $\mu_{OV,hoch} = 1,0$	33,33 % $\mu_{FT,gering} = 0,33$ $\mu_{FT,mittel} = 0,67$ $\mu_{FT,hoch} = 0,0$	33,33 % $\mu_{UE,gering} = 0,33$ $\mu_{UE,mittel} = 0,67$ $\mu_{UE,hoch} = 0,0$	$g - h: 0,047$ $g - q: 0,061$ $h - q: 0,0$
8. 61,53	<i>Trocknen von Holz</i>	0,2 $\mu_{SV,gering} = 0,0$ $\mu_{SV,mittel} = 0,78$ $\mu_{SV,hoch} = 0,22$	0,458 $\mu_{FV,gering} = 0,0$ $\mu_{FV,mittel} = 0,77$ $\mu_{FV,hoch} = 0,23$	0,313 $\mu_{OV,gering} = 0,0$ $\mu_{OV,mittel} = 0,0$ $\mu_{OV,hoch} = 1,0$	100 % $\mu_{FT,gering} = 0,0$ $\mu_{FT,mittel} = 0,0$ $\mu_{FT,hoch} = 1,0$	33,33 % $\mu_{UE,gering} = 0,33$ $\mu_{UE,mittel} = 0,67$ $\mu_{UE,hoch} = 0,0$	$g - h: 0,405$ $g - q: 0,385$ $h - q: 0,545$
9. 56,57	<i>Schwindverformungen</i>	0,301 $\mu_{SV,gering} = 0,0$ $\mu_{SV,mittel} = 0,0$ $\mu_{SV,hoch} = 1,0$	0,5 $\mu_{FV,gering} = 0,0$ $\mu_{FV,mittel} = 0,59$ $\mu_{FV,hoch} = 0,41$	0,141 $\mu_{OV,gering} = 0,0$ $\mu_{OV,mittel} = 0,9$ $\mu_{OV,hoch} = 0,1$	66,67 % $\mu_{FT,gering} = 0,0$ $\mu_{FT,mittel} = 0,0$ $\mu_{FT,hoch} = 1,0$	0 % $\mu_{UE,gering} = 1,0$ $\mu_{UE,mittel} = 0,0$ $\mu_{UE,hoch} = 0,0$	$g - h: 0,192$ $g - q: 0,125$ $h - q: 0,636$
10.* 53,39	<i>Erklärung der Kurzzeichen und Fachbegriffe (Holz)</i>	0,129 $\mu_{SV,gering} = 0,37$ $\mu_{SV,mittel} = 0,63$ $\mu_{SV,hoch} = 0,0$	0,48 $\mu_{FV,gering} = 0,0$ $\mu_{FV,mittel} = 0,68$ $\mu_{FV,hoch} = 0,32$	0,129 $\mu_{OV,gering} = 0,0$ $\mu_{OV,mittel} = 0,99$ $\mu_{OV,hoch} = 0,01$	66,67 % $\mu_{FT,gering} = 0,0$ $\mu_{FT,mittel} = 0,0$ $\mu_{FT,hoch} = 1,0$	33,33 % $\mu_{UE,gering} = 0,33$ $\mu_{UE,mittel} = 0,67$ $\mu_{UE,hoch} = 0,0$	$g - h: 0,2$ $g - q: 0,159$ $h - q: 0,170$
11. 52,21	<i>Relevante Eigensch. der für Bauzwecke eingesetzten Holzarten</i>	0,176 $\mu_{SV,gering} = 0,0$ $\mu_{SV,mittel} = 0,98$ $\mu_{SV,hoch} = 0,02$	0,429 $\mu_{FV,gering} = 0,0$ $\mu_{FV,mittel} = 0,91$ $\mu_{FV,hoch} = 0,09$	0,253 $\mu_{OV,gering} = 0,0$ $\mu_{OV,mittel} = 0,0$ $\mu_{OV,hoch} = 1,0$	66,67 % $\mu_{FT,gering} = 0,0$ $\mu_{FT,mittel} = 0,0$ $\mu_{FT,hoch} = 1,0$	33,33 % $\mu_{UE,gering} = 0,33$ $\mu_{UE,mittel} = 0,67$ $\mu_{UE,hoch} = 0,0$	$g - h: 0,190$ $g - q: 0,125$ $h - q: 0,636$
20.** 38,94	<i>Einfluss der Temperatur auf das Verformungsmodul von Holz</i>	0,238 $\mu_{SV,gering} = 0,0$ $\mu_{SV,mittel} = 0,48$ $\mu_{SV,hoch} = 0,52$	0,4 $\mu_{FV,gering} = 0,03$ $\mu_{FV,mittel} = 0,97$ $\mu_{FV,hoch} = 1,0$	0,149 $\mu_{OV,gering} = 0,0$ $\mu_{OV,mittel} = 0,82$ $\mu_{OV,hoch} = 0,18$	33,33 % $\mu_{FT,gering} = 0,33$ $\mu_{FT,mittel} = 0,67$ $\mu_{FT,hoch} = 0,0$	33,33 % $\mu_{UE,gering} = 0,33$ $\mu_{UE,mittel} = 0,67$ $\mu_{UE,hoch} = 0,0$	$g - h: 0,55$ $g - q: 0,097$ $h - q: 0,175$
*) Wissensseinheiten, die nachträglich aus dem Lehrpfad entfernt werden							
**) Wissensseinheit, die nachträglich in den Lehrpfad aufgenommen wird							

In der Tabelle dargestellt werden alle Wissensseinheiten der obersten Rangplätze im Umfang von L = 11 Seiten. Wissensseinheiten, die vom Programm später entfernt wurden, sind gekennzeichnet. Ebenfalls in der Tabelle enthalten ist die Seite, die durch die Berücksichtigung der Unterthemen später hinzukam.

Es wurden zwei Seiten aus dem Bereich der 11 besten Seiten ausgeschlossen, weil sie zu geringe Ähnlichkeiten mit den einzelnen Unterthemen aufwiesen. Dies waren die Seiten „Stromquellen“ und „Erklärung der Kurzzeichen und Fachbegriffe (Holz)“. Zunächst muss erläutert werden, wieso diese später ausgeschlossenen Seiten in der Relevanzberechnung überhaupt so hoch eingeordnet wurden, obwohl sie thematisch nicht zum Lehrpfad passen. Bei der Seite „Erklärung der Kurzzeichen und Fachbegriffe (Holz)“ ist dies offensichtlich: Diese Wissensseinheit definiert neben mehreren anderen Fachbegriffen auch das „Stehvermögen“, das in den Kontext des „Quellens“ gehört. Bei der Berechnung ergaben sich entsprechend hohe Ähnlichkeitswerte. Die Seite „Stromquellen“ beinhaltet Eigenschaften von Stromquellen für das Schweißen, hat also mit der Gesamthematik

der Suchanfrage objektiv nichts zu tun. Bei dem Wort „Quellen“ handelt es sich um ein Homonym. Es gibt für dieses Wort also mehrere Kontexte, die jedoch nicht getrennt berücksichtigt werden. Da in den Inhalten des WiBA-Net an mehreren Stellen elektrischer Strom thematisiert wird, weist der Wortstamm „quell“ als Konstituent von „Stromquelle“ in seinem globalen Kontext auch Begriffe aus dieser Thematik auf. Diese wurden bei der Ähnlichkeitsberechnung mit herangezogen, was die ausreichend hohe Ähnlichkeit für den 7. Rangplatz erklärt.

Die nachträglich in den Lehrpfad aufgenommene Seite „Einfluss der Temperatur auf das Verformungsmodul von Holz“ thematisiert neben dem Temperatureinfluss auch die Holzfeuchte, die als Parameter bei der Bewertung des Temperatureinflusses zu berücksichtigen ist.

Damit eine Seite aus einem Lehrpfad ausgeschlossen wird, muss sie eine hinreichend niedrige Ähnlichkeit mit den Unterthemen aufweisen. Die in Tabelle 6.4 angegebenen Ähnlichkeiten $\text{sim}_{\text{UT},n}$ sind insbesondere bei der Wissensseinheit „Stromquellen“ im Vergleich um fast das zehnfache kleiner als bei den anderen Wissensseinheiten. Dieser Unterschied ist bei der Wissensseinheit „Erklärung der Kurzzeichen und Fachbegriffe (Holz)“ nicht ganz so sehr ausgeprägt, jedoch fällt bei dieser auf, dass alle $\text{sim}_{\text{UT},n}$ -Werte im untersten Bereich angesiedelt sind, ihr Mittelwert beträgt 0,176, weniger als die Hälfte von 0,356, dem Mittelwert der Wissensseinheiten, die im Lehrpfad verbleiben. Demgegenüber zeichnet sich die nachträglich in den Lehrpfad aufgenommene Wissensseinheit durch einen sehr hohen $\text{sim}_{\text{UT},n}$ -Wert für die Suchwortkombination „Gleichgewichtsfeuchte“ – „Holz“ aus.

Um berechnen zu können, welche Wissensseinheiten ausgeschlossen oder aufgenommen werden, wurden die Ähnlichkeiten der Wissensseinheiten $\text{sim}_{\text{UT},n}$ mit den einzelnen Unterthemen jeweils in eine Rangfolge, beginnend mit der größten Ähnlichkeit, gebracht. Aus den drei Rangfolgen, dargestellt in Tabelle 6.5, wurde für jede Wissensseinheit nach Formel (5.9) der Wert $\text{sim}_{\text{UT}}(d_i)$ berechnet.

Tabelle 6.5: Zusammenstellung der ersten L + 20 Wissensseinheiten, nach der Ähnlichkeit mit den drei Suchwortkombinationen (Angabe in der Klammer) sortiert

Gleichgewichtsfeuchte – Holz	Holz – Quellen	Gleichgewichtsfeuchte – Quellen
1. Gleichgewichtsfeuchte von Holz (0.6)	1. Schwindverformungen (0.64)	1. Gleichgewichtsfeuchte von Holz (0.59)
2. <u>Einfluss der Temperatur auf das Verformungsmodul von Holz</u> (0.55)	2. Relevante Eigenschaften der für Bauzwecke eingesetzten Holzarten (0.64)	2. Trocknen von Holz (0.38)
3. Sorptionsisotherme und Mechanismen der Feuchteaufnahme (0.54)	3. Trocknen von Holz (0.55)	3. Sorptionsisotherme und Mechanismen der Feuchteaufnahme (0.36)
4. Trocknen von Holz (0.4)	4. Quell- und Schwindmaße von Holz (0.5)	4. Quell- und Schwindmaße von Holz (0.33)
5. Einfluss der Holzfeuchte auf das Verformungsverhalten (0.4)	5. Schwinden und Quellen von Holz (0.45)	5. Schwinden und Quellen von Holz (0.31)
6. Gefährdungsklasse 1 (0.39)	6. Gleichgewichtsfeuchte von Holz (0.45)	6. Charakteristische Eigenschaften – Schwinden und Quellen (0.25)
7. Schwinden und Quellen von Holz (0.35)	7. Charakteristische Eigenschaften – Schwinden und Quellen (0.4)	7. Zusammenstellung der technischen Kennwerte von Nadelhölzern (0.23)
8. Der Werkstoff Holz im Bauwesen (0.29)	8. Der Werkstoff Holz im Bauwesen (0.36)	8. Zeitschiene der Entwicklung des Holzbaus (0.19)
9. Zusammenstellung der technischen Kennwerte von Nadelhölzern (0.29)	9. Sorptionsisotherme und Mechanismen der Feuchteaufnahme (0.35)	9. Einfluss der Holzfeuchte auf das Verformungsverhalten (0.18)
		10. Gefährdungsklasse 1 (0.16)

Gleichgewichtsfeuchte – Holz	Holz – Quellen	Gleichgewichtsfeuchte – Quellen
10. Gefährdungsklasse 0 (0.29) 11. Charakteristische Eigenschaften – Feuchteverhalten (0.24) 12. Maßnahmen des Holzschutzes (0.2) 13. Charakteristische Eigenschaften – Schwinden und Quellen (0.2) 14. Erklärung der Kurzzeichen und Fachbegriffe (Holz) (0.2) 15. Erklärung der Kurzzeichen und Fachbegriffe (Holz) (0.2)* 16. Gefährdungsklasse 2 (0.2) 17. Schwindverformungen (0.19) 18. Relevante Eigenschaften der für Bauzwecke eingesetzten Holzarten (0.19) 19. Grundprinzipien des Konstruktiven Holzschutzes – Begriffsklärung (0.19) 20. Quell- und Schwindmaße von Holz (0.19) 21. Zeitschiene der Entwicklung des Holzbau (0.18) 22. Konstruktiver Holzschutz – feuchtegefährdete Holzbauteile (0.15) 23. Konstruktiver Holzschutz – Gefährdungsklassen (GK) (0.14) 24. Konstruktiver Holzschutz – Schutz gegen Feuchteleitung (0.14) 25. Sinnlichkeit – Geruch (0.13) 26. Holzskelettbau – Wandschnitte (0.1) 27. Metall-Lichtbogen-Schweißen (0.09) 28. Anwendungsbeispiel – Holzfassaden (0.09) 29. Beurteilen und Bewerten von Holzschutzmitteln (0.09) 30. Holzschutzmittel-Einbringverfahren (0.09) 31. Stromquellen (0.05)	10. Zusammenstellung der technischen Kennwerte von Nadelhölzern (0.32) 11. Gefährdungsklasse 2 (0.32) 12. Gefährdungsklasse 1 (0.31) 13. Anwendungsbeispiel – Holzfassaden (0.27) 14. Holzschutzmittel-Einbringverfahren (0.22) 15. Holzskelettbau – Wandschnitte (0.22) 16. Gefährdungsklasse 0 (0.22) 17. Maßnahmen des Holzschutzes (0.21) 18. Zeitschiene der Entwicklung des Holzbau (0.19) 19. Einfluss der Holzfeuchte auf das Verformungsverhalten (0.18) 20. Grundprinzipien des Konstruktiven Holzschutzes – Begriffsklärung (0.18) 21. Konstruktiver Holzschutz – Gefährdungsklassen (GK) (0.18) 22. <u>Einfluss der Temperatur auf das Verformungsmodul von Holz (0.17)</u> 23. Konstruktiver Holzschutz – Schutz gegen Feuchteleitung (0.17) 24. Erklärung der Kurzzeichen und Fachbegriffe (Holz) (0.17) 25. Erklärung der Kurzzeichen und Fachbegriffe (Holz) (0.17)* 26. Charakteristische Eigenschaften – Feuchteverhalten (0.13) 27. Konstruktiver Holzschutz – feuchtegefährdete Holzbauteile (0.12) 28. Sinnlichkeit – Geruch (0.12) 29. Beurteilen und Bewerten von Holzschutzmitteln (0.08) 30. Metall-Lichtbogen-Schweißen (0.04) 31. Stromquellen (0.0)	11. Erklärung der Kurzzeichen und Fachbegriffe (Holz) (0.16) 12. Erklärung der Kurzzeichen und Fachbegriffe (Holz) (0.16)* 13. Maßnahmen des Holzschutzes (0.15) 14. Der Werkstoff Holz im Bauwesen (0.13) 15. Gefährdungsklasse 0 (0.13) 16. Schwindverformungen (0.13) 17. Relevante Eigenschaften der für Bauzwecke eingesetzten Holzarten (0.13) 18. Charakteristische Eigenschaften – Feuchteverhalten (0.12) 19. Konstruktiver Holzschutz – Schutz gegen Feuchteleitung (0.12) 20. Sinnlichkeit – Geruch (0.1) 21. Konstruktiver Holzschutz – Gefährdungsklassen (GK) (0.1) 22. Grundprinzipien des Konstruktiven Holzschutzes – Begriffsklärung (0.1) 23. <u>Einfluss der Temperatur auf das Verformungsmodul von Holz (0.1)</u> 24. Holzskelettbau – Wandschnitte (0.09) 25. Gefährdungsklasse 2 (0.07) 26. Metall-Lichtbogen-Schweißen (0.07) 27. Konstruktiver Holzschutz – feuchtegefährdete Holzbauteile (0.06) 28. Stromquellen (0.06) 29. Anwendungsbeispiel – Holzfassaden (0.06) 30. Beurteilen und Bewerten von Holzschutzmitteln (0.06) 31. Holzschutzmittel-Einbringverfahren (0.04)
<u>unterstrichen</u> : Wissensseinheit, die nachträglich aufgenommen wird fett : Wissensseinheiten, die nachträglich entfernt werden *) Eine Wissensseinheit dieses Titels ist wegen einer nachträglichen Korrektur doppelt im System vorhanden.		

Für die Wissensseinheit „Stromquellen“ galt z.B.:

$$sim_{UT}('Stromquellen') = \frac{\frac{0,05}{31} + \frac{0,0}{31} + \frac{0,06}{28}}{3} = 0,0013 \quad (6.1)$$

Analog ergab sich für die Wissensseinheit „Erklärung der Kurzzeichen und Fachbegriffe (Holz)“ ein sim_{UT} -Wert von 0,0110. Diese Werte stellen die niedrigsten sim_{UT} -Werte der im Lehrpfad vorhandenen Wissensseinheiten dar. Die Auswertung fand anhand der sechs Wissensseinheiten⁷³ mit den niedrigsten sim_{UT} -Werten statt. Diese sind in Tabelle 6.6 verzeichnet. Die durchschnittliche Differenz der sechs niedrigsten sim_{UT} -Werte betrug 0,0142, der nächsthöhere sim_{UT} -Wert betrug 0,0363. Die Differenz Δsim_{UT} zwischen dem zweit- und dem drittletzten Wert betrug 0,0252 – 69,31% des drittletzten Wertes. Die Differenz an dieser Stelle war 1,77 mal so groß wie die durch-

⁷³ als Bezugswert

schnittliche Differenz. Damit lagen die in Abschnitt 5.3.5 definierten Abbruchkriterien vor, es wurden also die zwei Wissensseinheiten mit den niedrigsten sim_{UT} -Werten abgetrennt.

Tabelle 6.6: Kennwerte zur Bestimmung einer nachträglichen Aufnahme von Wissensseinheiten

Rang	Wissenseinheit	$\text{sim}_{\text{UT}}(\text{d}_i)$	$\Delta\text{sim}_{\text{UT}}$	bez. $\Delta\text{sim}_{\text{UT}}$	$\frac{\Delta\text{sim}_{\text{UT}}}{\Delta\text{sim}_{\text{UT},m}}$	Abbruch-kriterium
1	Stromquellen	0,0013	0,0097	88,18%	0,6631	erfüllt
2	Erklärung der Kurzzeichen und Fachbegriffe (Holz)	0,0110				
3	Einfluss der Holzfeuchte auf das Verformungsverhalten	0,0363	0,0252	69,31%	1,7746	
			0,0018	4,62%	0,1268	
4	Charakteristische Eigenschaften - Schwinden und Quellen	0,0381	0,0297	43,82%	2,0915	
			0,0048	6,64%	0,3380	
5	Schwinden und Quellen von Holz	0,0678				
6	Quell- und Schwindmaße von Holz	0,0726				
Mittelwert $\Delta\text{sim}_{\text{UT},m} =$			0,0142			

Für die Wissensseinheit „Einfluss der Temperatur auf das Verformungsmodul von Holz“ berechnete das Programm analog einen sim_{UT} -Wert von 0,0957. Dies stellte den höchsten sim_{UT} -Wert der potentiell in den Lehrpfad aufzunehmenden Wissensseinheiten dar. Die Differenz zum zweithöchsten Wert stellte mit 0,0599 das Maximum der Differenzen zwischen den höchsten sechs sim_{UT} -Werten dar. Aus diesem Grund wurde nur diese eine Seite zusätzlich in den Lehrpfad aufgenommen.

Nachdem alle im Lehrpfad vorhandenen Wissensseinheiten bestimmt waren, wurden sie zu Clustern zusammengefasst. Dazu wurde zunächst die Ähnlichkeit zwischen den Dokumentvektoren der verschiedenen Wissensseinheiten als Cosinusmaß bestimmt. Im Single-Pass-Cluster-Verfahren werden Abstände zwischen Dokumenten betrachtet, sodass kleinere Werte für größere Ähnlichkeiten stehen. Um ein solches Verhalten auch für die mit dem Cosinusmaß bestimmten Werte zu erhalten, wurden sie in diesem Fall nicht direkt als Ähnlichkeitsmaß verwendet, sondern ihre jeweilige Differenz zu 1, das Clusterdifferenzmaß $1 - \cos(\alpha)$.

Die jeweiligen Clusterdifferenzmaße sind in Tabelle 6.7 zusammengestellt. Aus den Clusterdifferenzmaßen jeweils einer Spalte lässt sich ein Mittelwert $\text{MW}_{\text{Spalte}}$ berechnen, der als mittleres Clusterdifferenzmaß einer Wissensseinheit zu allen übrigen Wissensseinheiten interpretiert werden kann. Der Mittelwert aller Clusterdifferenzmaße betrug 0,797, die Standardabweichung 0,079. Daraus ergab sich ein Grenzwert für die Clusterzugehörigkeit von $G_c = 0,797 - 0,079 = 0,718$. In der Zeile Δ_{SPV} ist das Clusterdifferenzmaß jeder Wissensseinheit zum gemeinsamen Schwerpunktsvektor

SPV verzeichnet. Die Wissensseinheit, die den höchsten Relevanzwert R und bei mehreren Seiten mit demselben Relevanzwert gleichzeitig die größte Ähnlichkeit mit dem gemeinsamen Schwerpunktsvektor aller Wissensseinheiten aufweist, wird als Kern des ersten Clusters ausgewählt. Dies war in diesem Fall die Seite „Gleichgewichtsfeuchte von Holz“ mit $R = 89,17$ und $\Delta_{SPV} = 0,45$.

Tabelle 6.7: Ähnlichkeitsmaße zwischen den Wissensseinheiten des Lehrpfads

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1)		0,79	0,73	0,84	0,83	0,52	0,79	0,82	0,84	0,86
(2)	0,79		0,73	0,86	0,84	0,79	0,86	0,83	0,88	0,91
(3)	0,73	0,73		0,79	0,80	0,62	0,71	0,75	0,85	0,87
(4)	0,84	0,86	0,79		0,74	0,86	0,64	0,75	0,77	0,78
(5)	0,83	0,84	0,80	0,74		0,87	0,84	0,78	0,78	0,82
(6)	0,52	0,79	0,62	0,86	0,87		0,75	0,83	0,87	0,89
(7)	0,79	0,86	0,71	0,64	0,84	0,75		0,76	0,83	0,88
(8)	0,82	0,83	0,75	0,75	0,78	0,83	0,76		0,88	0,81
(9)	0,84	0,88	0,85	0,77	0,78	0,87	0,83	0,88		0,63
(10)	0,86	0,91	0,87	0,78	0,82	0,89	0,88	0,81	0,63	
MW_{Spalte}	0,78	0,74	0,68	0,69	0,72	0,72	0,70	0,71	0,72	0,73
Δ_{SPV}	0,49	0,41	0,38	0,45	0,44	0,39	0,32	0,48	0,30	0,33
(1) Relevante Eigenschaften der für Bauzwecke eingesetzten Holzarten (2) Charakteristische Eigenschaften – Schwinden und Quellen (3) Quell- und Schwindmaße von Holz (4) Gleichgewichtsfeuchte von Holz (5) Sorptionsisotherme und Mechanismen der Feuchtaufnahme (6) Schwindverformungen (7) Schwinden und Quellen von Holz (8) Trocknen von Holz (9) Einfluss der Holzfeuchte auf das Verformungsverhalten (10) Einfluss der Temperatur auf das Verformungsmodul von Holz										

Nach Formel (5.11) ändert sich G_c je nach Größe des aktuellen Clusters. Die Zuordnung der Wissensseinheiten zu Clustern ist in Tabelle 6.8 dargestellt. Nach dem Zufügen einer Wissensseinheit zu einem Cluster wird der Schwerpunkt des Clusters neu berechnet. Wird ein neues Cluster angelegt, so ist dessen Kern nicht die letzte Seite, die sich an das vorige Cluster nicht mehr anfügen lässt, sondern die Seite, die das größte Clusterdifferenzmaß zum aktuellen Cluster aufweist, diesem also am unähnlichsten ist. Sowohl die Reihenfolge der Cluster als auch die Untergliederung der Cluster in einen Clusterkern und zugeordnete Seiten ist zu diesem Zeitpunkt noch irrelevant.

Tabelle 6.8: Ablauf der Clusterbildung des Lehrpfads

Nr.	Clusterkern	$G_{C,n}$	untersuchte Seite	Ähnlichkeitsmaß zum Clusterkern	Aufnahme in Cluster?
1	Gleichgewichtsfeuchte von Holz	0,86	Schwinden und Quellen von Holz	0,64	ja
		0,72	Trocknen von Holz	0,75	nein
2	Charakteristische Eigenschaften – Schwinden und Quellen	0,86	Quell- und Schwindmaße von Holz	0,74	ja
		0,72	Schwindverformungen	0,68	ja
		0,68	Relevante Eigenschaften der für Bauzwecke eingesetzten Holzarten	0,62	ja
		0,64	Trocknen von Holz	0,78	nein
3	Einfluss der Temperatur auf das Verformungsmodul von Holz	0,86	Einfluss der Holzfeuchte auf das Verformungsverhalten	0,63	ja
		0,72	Sorptionsisotherme und Mechanismen der Feuchteaufnahme	0,78	nein
4	Trocknen von Holz	0,86	Sorptionsisotherme und Mechanismen der Feuchteaufnahme	0,79	ja

Im nächsten Schritt wurden die Cluster des Lehrpfads nach dem in Abschnitt 5.4.3 beschriebenen Verfahren in eine angemessene didaktische Reihenfolge gebracht. Dazu wurden zunächst die Parameter ermittelt, die für die Zuordnung der Cluster zu den didaktischen Kategorien benötigt werden. Die Parameter können Tabelle 6.9 entnommen werden. Da noch keine Haupt- und Nebenseiten bestimmt wurden, werden die Cluster in dieser Tabelle anhand der oben beschriebenen Clusterkerne identifiziert. Die Parameter für die Zugehörigkeitsfunktionen der Eingangsgrößen können bis auf den Anteil der Determinativkomposita und die über das Cluster gemittelten Ähnlichkeiten mit den ursprünglichen Suchbegriffen Tabelle 5.8 entnommen werden. Für den Anteil an Determinativkomposita ergab sich $E_{\text{oben}} = 0,08$, $E_{\text{unten}} = 0,0$ und $E_m = (0,08 + 0,03) / 4 = 0,03$. Für die Ähnlichkeiten $\text{sim}_{\text{ov}}(d_i)$ wurden die Parameter aus Abbildung 6.3 übernommen.

Nach Durchführung der Inferenz und der Defuzzifizierung ergaben sich die angegebenen scharfen Ausgangswerte für „Konkretheit“, „Komplexität“ und „Spezialisierung“. Auch ohne Rechnung sieht man leicht, dass die Anordnung vom Allgemeinen zum Speziellen am breitesten verteilt ist und deshalb als Lehrpfadsortierung auszuwählen ist. Dies wurde durch die für diese Anordnung maximale Standardabweichung bestätigt. Die Mittelwerte MW und Standardabweichungen s sind in der Tabelle verzeichnet.

Tabelle 6.9: Bestimmung der Reihenfolge der Cluster

Nr.	1	2	3	4
Clusterkern	Gleichgewichtsfeuchte von Holz	Charakteristische Eigenschaften – Schwinden und Quellen	Einfluss der Temperatur auf das Verformungsmodul von Holz	Trocknen von Holz
Anzahl Bilder / Animationen	1 / 0	7 / 1	1 / 1	3 / 0
sim_{ov}(d_i), gemittelt	0,3446	0,2071	0,1916	0,3074
Textlänge pro enthaltener Wissenseinheit	82	89,25	62	117,5
Anteil seltener Begriffe*	0,02	0,08	0,05	0,09
Anteil an Begriffen mit wenigen Kookkurrenten*	0,15	0,20	0,10	0,17
Anzahl / Anteil an Determinativkomposita*	0 / 0,0	7 / 0,08	2 / 0,03	0 / 0
„Konkretheit“ MW = 65,28, s = 15,28	80,56	50,00	50,00	80,56
„Komplexität“ MW = 33,11, s = 6,21	23,75	36,54	31,75	40,39
„Spezialisierung“ MW = 45,95, s = 18,70	34,75	77,88	31,19	40,01
*) bezogen auf die Gesamtextlänge des Clusters				

Für jedes Cluster wurde abschließend eine Hauptseite bestimmt, die übrigen Seiten des Clusters bildeten die Nebenseiten. Hierzu wurden fünf Parameter benötigt. Diese sind in Tabelle 6.10 angegeben.

Tabelle 6.10: Bestimmung von Haupt- und Nebenseiten

Nr.	Seite	Abstand	allgemein	A _{FT}	A _{UE}	sim _{ov}	H _i -Wert
1	Gleichgewichtsfeuchte von Holz	1,000	1,000	1,00	0,67	0,400	2,689
	Schwinden und Quellen von Holz	1,000	0,309	1,00	0,67	0,289	1,961
2	Charakteristische Eigenschaften – Schwinden und Quellen	0,837	1,000	0,67	0,33	0,191	2,234
	Quell- und Schwindmaße von Holz	0,928	0,385	0,67	0,67	0,242	1,838
	Relevante Eigenschaften der für Bauzwecke eingesetzten Holzarten	0,946	0,283	0,67	0,33	0,253	1,647
	Schwindverformungen	1,000	0,276	0,67	0,00	0,141	1,545
3	Einfluss der Temperatur auf das Verformungsmodul von Holz	1,000	1,000	0,33	0,33	0,149	2,272
	Einfluss der Holzfeuchte auf das Verformungsverhalten	1,000	0,534	0,33	0,33	0,234	1,834

Nr.	Seite	Abstand	allgemein	A _{FT}	A _{UE}	sim _{OV}	H _i -Wert
4	Trocknen von Holz	1,000	1,000	1,00	0,33	0,313	2,549
	Sorptionsisotherme und Mechanismen der Feuchteaufnahme	1,000	0,484	0,67	0,00	0,302	1,807

Die erste Spalte gibt den normierten Kehrwert des mittleren Abstands der jeweiligen Wissenseinheit von den übrigen Wissenseinheiten nach Formel (5.12) an – dieser ist bei Clustern mit maximal zwei Seiten immer 1,0. Die Seite mit dem geringsten Abstand zu den anderen Seiten – bzw. mit dessen höchstem Kehrwert – wird als thematisch zentral interpretiert. Die Kehrwerte der mittleren Abstände werden auf den Kehrwert des geringsten Abstandes bezogen, sodass dieser den Wert 1,0 erhält. Die zweite Spalte gibt die „Allgemeinverbindlichkeit“ einer Seite, bzw. nach Formel (5.13) den normierten Kehrwert ihrer „Spezialisierung“ an. Diese wurde wie bei der Bestimmung der Reihenfolge der Cluster ermittelt, indem die Begriffe mit weniger als 100 Kookkurrenten pro Textlänge gezählt wurden. Die Parameter A_{FT}, A_{UE} und sim_{OV} wurden aus der Relevanzberechnung (vgl. Tabelle 6.4) übernommen. Der Hierarchiewert H_i wurde mit Formel (5.14) berechnet, die letzten drei Parameter gingen dabei zu je einem Drittel in den Wert ein.

Als weiterführende Seiten fanden sich zur Wissenseinheit „Einfluss der Holzfeuchte auf das Verformungsverhalten“ (R = 72,78, entspricht 81,63 % des maximalen Relevanzwerts) die Seite „Verformungsverhalten von Holz unter Kurzzeitbelastung“ (sim_{cos}(d₁,d₂) = 0,3494) und zu der Wissenseinheit „Gleichgewichtsfeuchte von Holz“ (R = 89,17, entspricht 100 %) die Seiten „Temperaturabhängigkeit der Holzfestigkeit“ (sim_{cos}(d₂,d₃) = 0,2695) und „Abhängigkeit der Festigkeit von der Rohdichte“ (sim_{cos}(d₂,d₄) = 0,2545).

Die Bestimmung der Inhalte und der Struktur des Lehrpfads zu den Suchbegriffen „Gleichgewichtsfeuchte“, „Holz“ und „Quellen“ ist damit vollständig. Der Lehrpfad bietet für den Lerner einen vollständigen Überblick über das Schwind- und Quellverhalten von Holz und berücksichtigt dabei auch die Gleichgewichtsfeuchte.

Zum Vergleich soll ferner eine Modifikation des Lehrpfads betrachtet werden. Dazu wurde in der Suchanfrage der Begriff „Holz“ weggelassen. Es resultierten Veränderungen im Aufbau und im Inhalt des Lehrpfads. Der modifizierte Lehrpfad ist in Tabelle 6.11 dargestellt. Der offensichtlichste Unterschied ist ein Cluster bestehend aus den Seiten „Metall-Lichtbogen-Schweißen“ und „Stromquellen“. Die Kennwerte der Relevanzberechnung für diese Seiten sind in Tabelle 6.12 angegeben. Diese Thematik wurde im ursprünglichen Lehrpfad durch den Bezug auf den Baustoff Holz unterdrückt. Es soll hier untersucht werden, wie die Seiten in der Modifikation in den Lehrpfad gelangten.

Tabelle 6.11: Inhalte und Struktur des modifizierten Lehrpfads 1

Suchbegriffe / Kenngrößen	Struktur des erzeugten Lehrpfads
<ul style="list-style-type: none"> ● Gleichgewichtsfeuchte ● Quellen <p>Sollwert L: 11 Seiten</p> <p>tatsächlicher Umfang: 12 Seiten</p> <p>Ordnungsprinzip: vom Allgemeinen zum Speziellen</p>	<p>► Einfluss der Holzfeuchte auf das Verformungsverhalten</p> <ul style="list-style-type: none"> - Lastunabhängige Verformung (Schwinden und Quellen) <p>► Charakteristische Eigenschaften – Feuchteverhalten</p> <ul style="list-style-type: none"> - Charakteristische Eigenschaften – Schwinden und Quelle - Trocknen von Holz <p>► Lösung für idealplastisches Werkstoffgesetz</p> <p>► Quell- und Schwindmaße von Holz</p> <ul style="list-style-type: none"> - <i>Schwindverformungen</i> - Gleichgewichtsfeuchte von Holz - Schwinden und Quellen von Holz - Sorptionsisotherme und Mechanismen der Feuchtaufnahme <p>► Metall-Lichtbogen-Schweißen</p> <ul style="list-style-type: none"> - Stromquellen
Kennzeichnung der Strukturierung: ► Hauptseiten, - Nebenseiten, - <i>weiterführende Seiten</i>	

Da bei dieser Suchanfrage nur zwei Suchbegriffe vorkamen, wurden als Unterthemen keine Ähnlichkeiten zu den Kombinationen der Suchbegriffe untersucht, sondern nur die zu den einzelnen Suchbegriffen selber („g“: „Gleichgewichtsfeuchte“, „q“: „Quellen“). Das Unterthema „Gleichgewichtsfeuchte“ wies zu den beiden betrachteten Wissensseinheiten eine sehr geringe Ähnlichkeit auf. Es fällt bei der Wissensseinheit „Stromquellen“ auf, dass der Wert $\text{sim}_{\text{FV}}(d_i)$ derselbe wie ohne Modifikation der Suchanfrage ist (vgl. Tabelle 6.4). Das liegt daran, dass diese Wissensseinheit in beiden Fällen durch dieselben Suchbegriffe gefunden wurde, deren Kookkurrenten zur Berechnung des Ähnlichkeitsmaßes $\text{sim}_{\text{FV}}(d_i)$ genutzt wurden. Allerdings unterscheiden sich die entsprechenden Zugehörigkeitsgrade der Eingangswerte, was durch die unterschiedlichen Parameter der Zugehörigkeitsfunktionen begründet ist. Die Ähnlichkeit mit den Original-Suchbegriffen ist hoch. Dies kann auch dann der Fall sein, wenn nur ein Original-Suchbegriff – bzw. Worte aus dessen zur Ähnlichkeitsberechnung verwendetem Kontext – in der Wissensseinheit vorkommen.

Tabelle 6.12: Kennwerte der Relevanzberechnung der Wissensseinheiten für die Suchbegriffe „Gleichgewichtsfeuchte“ und „Quellen“ (Auszug)

Rang / Relevanz R	Titel	$\text{sim}_{\text{SV}}(d_i) / \mu_{\text{SV}}$	$\text{sim}_{\text{FV}}(d_i) / \mu_{\text{FV}}$	$\text{sim}_{\text{OV}}(d_i) / \mu_{\text{OV}}$	$A_{\text{FT}} / \mu_{\text{FT}}$	$A_{\text{UE}} / \mu_{\text{UE}}$	$\text{sim}_{\text{UT},n}(d_i)$
5. 70,26	<i>Stromquellen</i>	0,2 $\mu_{\text{SV,gering}} = 0,0$ $\mu_{\text{SV,mittel}} = 0,31$ $\mu_{\text{SV,hoch}} = 0,69$	0,486 $\mu_{\text{FV,gering}} = 0,0$ $\mu_{\text{FV,mittel}} = 0,69$ $\mu_{\text{FV,hoch}} = 0,31$	0,388 $\mu_{\text{OV,gering}} = 0,0$ $\mu_{\text{OV,mittel}} = 0,0$ $\mu_{\text{OV,hoch}} = 1,0$	50 % $\mu_{\text{FT,gering}} = 0,00$ $\mu_{\text{FT,mittel}} = 0,5$ $\mu_{\text{FT,hoch}} = 0,5$	50 % $\mu_{\text{UE,gering}} = 0,0$ $\mu_{\text{UE,mittel}} = 0,5$ $\mu_{\text{UE,hoch}} = 0,5$	g: 0,021 q: 0,155
9. 50,03	<i>Metall-Lichtbogen-Schweißen</i>	0,122 $\mu_{\text{SV,gering}} = 0,12$ $\mu_{\text{SV,mittel}} = 0,88$ $\mu_{\text{SV,hoch}} = 0,0$	0,442 $\mu_{\text{FV,gering}} = 0,0$ $\mu_{\text{FV,mittel}} = 0,88$ $\mu_{\text{FV,hoch}} = 0,12$	0,253 $\mu_{\text{OV,gering}} = 0,0$ $\mu_{\text{OV,mittel}} = 0,0$ $\mu_{\text{OV,hoch}} = 1,0$	50 % $\mu_{\text{FT,gering}} = 0,0$ $\mu_{\text{FT,mittel}} = 0,5$ $\mu_{\text{FT,hoch}} = 0,5$	0 % $\mu_{\text{UE,gering}} = 1,0$ $\mu_{\text{UE,mittel}} = 0,0$ $\mu_{\text{UE,hoch}} = 0,0$	g: 0,01 q: 0,124

Die beiden betrachteten Wissenseinheiten wurden vom Programm in einem Cluster zusammengefasst. Sie wiesen ein Clusterdifferenzmaß von 0,79 zueinander auf. Das mittlere Clusterdifferenzmaß der Seite „Stromquellen“ zu den übrigen Seiten betrug 0,940, das mittlere Clusterdifferenzmaß der Seite „Metall-Lichtbogen-Schweißen“ 0,937. Alle Seiten hatten untereinander ein durchschnittliches Clusterdifferenzmaß von 0,874. Der von den übrigen Seiten des Lehrpfads, die sich hauptsächlich mit dem Quellen von Holz beschäftigen, völlig unterschiedliche Themenblock des elektrischen Stroms wurde also vom System auch rechnerisch als solcher erkannt und entsprechend in einem separaten Cluster angezeigt.

6.4 Lehrpfad 2

Der zweite Lehrpfad hatte die Frage nach zwei unterschiedlichen Herstellungsverfahren für Kunststoffe, nämlich Polykondensation und Polymerisation, zum Thema. Von einem fiktiven Lerner wurde angenommen, dass er eine Abgrenzung der beiden Begriffe sowie deren Einordnung in den Kontext der Kunststoffe benötigt. Der erstellte Lehrpfad besteht aus 11 Seiten, die vier Cluster bilden. Die Struktur des Lehrpfads ist in Tabelle 6.13 dargestellt.

Tabelle 6.13: Zusammenstellung der untersuchten Lehrpfade

Suchbegriffe / Kenngrößen	Struktur des erzeugten Lehrpfads
<ul style="list-style-type: none"> ● Polykondensation ● Polymerisation <p>Sollwert L: 12 Seiten</p> <p>tatsächlicher Umfang: 11 Seiten</p> <p>Ordnungsprinzip: vom Abstrakten zum Konkreten</p>	<ul style="list-style-type: none"> ► Polymerisation – radikalisch <ul style="list-style-type: none"> - Polymerisation – ionisch ► Verarbeiten der Duroplaste-Polykondensationsprodukte <ul style="list-style-type: none"> - Vergleich: Polymerisation-Polykondensation-Polyaddition - Polykondensation - Polyaddition ► Einteilung nach dem Herstellungsprinzip <ul style="list-style-type: none"> - <i>Kurzzeichen für Kunststoffe</i> - <i>Variation der Netzkettenlänge bei vernetzten Kunststoffen</i> - Reaktion vom Monomeren zum Polymeren - Kunststoffe ► Werkstoffe zur Verklebung mit Zweikomponentenklebern <ul style="list-style-type: none"> - Einführung – Was sind Kunststoffe?
Kennzeichnung der Strukturierung: ► Hauptseiten, - Nebenseiten, - <i>weiterführende Seiten</i>	

Bei den Begriffen „Polykondensation“ und „Polymerisation“ – sowie bei der „Polyaddition“ – handelt es sich um Kohyponyme. Dies sind Unterbegriffe eines gemeinsamen Oberbegriffs, die auf derselben taxonomischen Ebene angesiedelt sind. Sie bezeichnen analoge Vorgänge und werden im Rahmen der „Werkstoffe im Bauwesen“ in ähnlichen Kontexten gebraucht. Deshalb kann anhand

dieser Suchbegriffe bzw. des daraus erstellten Lehrpfads aufgezeigt werden, wie sich eine Abwandlung von Suchbegriffen auf die Inhalte und die Strukturierung von Inhalten auswirkt.

Untersucht wurden neben der Ausgangskombination „Polykondensation – Polymerisation“ auch die Kombinationen „Polyaddition – Polykondensation“, „Polyaddition – Polymerisation“ sowie „Polyaddition – Polykondensation – Polymerisation“. Die erstellten Lehrpfade inkl. ihrer Reihenfolge und Gliederung finden sich in Tabelle 6.14. Da weiterführende Seiten nicht direkt von der Suchanfrage, sondern nur von den im Lehrpfad vorhandenen Seiten abhängig sind, wurden sie in den Vergleich nicht mit einbezogen.

Tabelle 6.14: Vergleich der verschiedenen Alternativen zu Lehrpfad 2

Polykondensation – Polymerisation	Polyaddition – Polykondensation	Polyaddition – Polymerisation	Polyaddition – Polykondensation – Polymerisation
<ul style="list-style-type: none"> ► Polymerisation – radikalisch <ul style="list-style-type: none"> - Polymerisation – ionisch ► Verarbeiten der Duroplaste-Polykondensationsprodukte <ul style="list-style-type: none"> - Vergleich: Polymerisation-Polykondensation-Polyaddition - Polykondensation - Polyaddition ► Einteilung nach dem Herstellungsprinzip <ul style="list-style-type: none"> - Reaktion vom Monomeren zum Polymeren - Kunststoffe ► Werkstoffe zur Verklebung mit Zweikomponentenklebern <ul style="list-style-type: none"> - Einführung – Was sind Kunststoffe? 	<ul style="list-style-type: none"> ► <u>Polykondensation am Beispiel von Phenolharz</u> <ul style="list-style-type: none"> - Polyaddition - Verarbeiten der Duroplaste-Polykondensationsprodukte ► Einteilung nach dem Herstellungsprinzip <ul style="list-style-type: none"> - Reaktion vom Monomeren zum Polymeren ► Einführung – Was sind Kunststoffe? ► Werkstoffe zur Verklebung mit Zweikomponentenklebern <ul style="list-style-type: none"> - <u>Einkomponentenklebstoffe</u> ► Polykondensation <ul style="list-style-type: none"> - Vergleich: Polymerisation-Polykondensation-Polyaddition 	<ul style="list-style-type: none"> ► Polymerisation – radikalisch <ul style="list-style-type: none"> - Polymerisation – ionisch ► Verarbeiten der Duroplaste-Polykondensationsprodukte <ul style="list-style-type: none"> - Vergleich: Polymerisation-Polykondensation-Polyaddition - Polyaddition - Polykondensation ► Abbinden mit chemischer Reaktion – Reaktionskleber und ihre Umgebungsbedingungen <ul style="list-style-type: none"> - <u>Abbindevorgänge beim Kleben</u> ► Einteilung nach dem Herstellungsprinzip <ul style="list-style-type: none"> - Reaktion vom Monomeren zum Polymeren - Kunststoffe ► Einführung – Was sind Kunststoffe? 	<ul style="list-style-type: none"> ► <u>Zeitschiene der geschichtlichen Entwicklung von Kunststoffen</u> <ul style="list-style-type: none"> - Kunststoffe ► Abbinden mit chemischer Reaktion – Reaktionskleber und ihre Umgebungsbedingungen ► Verarbeiten der Duroplaste-Polykondensationsprodukte <ul style="list-style-type: none"> - Vergleich: Polymerisation-Polykondensation-Polyaddition - Polyaddition - Polykondensation ► Einteilung nach dem Herstellungsprinzip <ul style="list-style-type: none"> - Reaktion vom Monomeren zum Polymeren - <u>Kurzzeichen für Kunststoffe</u> ► Polymerisation – radikalisch <ul style="list-style-type: none"> - Polymerisation – ionisch
Ordnungsprinzip: abstrakt → konkret	Ordnungsprinzip: abstrakt → konkret	Ordnungsprinzip: abstrakt → konkret	Ordnungsprinzip: einfach → komplex
Kennzeichnung der Strukturierung: ► Hauptseiten, - Nebenseiten fett: Wissensseinheiten, die in den Lehrpfaden zu allen vier Suchwortkombinationen vorkommen <u>unterstrichen:</u> Wissensseinheiten, die nur im Lehrpfad zu einer Suchwortkombination vorkommen			

An diesen Lehrpfaden kann nachvollzogen werden, wie sich geringe Änderungen bei den Suchbegriffen im Inhalt und in der Struktur der Lehrpfade niederschlagen. Die Lehrpfade lassen deutlich

einen Bereich erkennen, der allen gemeinsam ist (in der Tabelle fett markiert) und Seiten, die jeweils nur bei einem Lehrpfad auftreten (in der Tabelle unterstrichen). Das Vorhandensein der gemeinsamen Basis zeigt, dass durch die Fuzzy-Berechnung als relevanteste Seiten zunächst einführende und übergreifende Themen ausgewählt wurden. Diese können dem Lerner als Grundlage dienen, wenn das reine Spezialwissen zu einer vertiefenden Suchanfrage ihn noch überfordert.

Die insgesamt relevanteste Wissensseinheit, die folglich als Startseite für die Clusterbildung diente, war in allen Fällen „Einteilung nach dem Herstellungsprinzip“. Da eine identische Startseite für die Clusterbildung bei allen Varianten vorlag, kann der Ablauf der Clusterbildung unter dem Einfluss unterschiedlicher zu clusternder Wissensseinheiten anschaulich nachvollzogen werden. Die Reihenfolge der Clusterbildung ist in Tabelle 6.15 dargestellt – diese ist nicht zu verwechseln mit der späteren Reihenfolge der Cluster im Lehrpfad.

In dieser Tabelle sind für die vier betrachteten Suchwortkombinationen die einzelnen Schritte bei der Clusterbildung angegeben⁷⁴. Die Zellen der Tabelle stellen die Cluster zu den entsprechenden Lehrpfaden dar. Ganz oben in jeder Tabellenzelle steht jeweils in fetter Schrift die Wissensseinheit, die den Clusterkern bildet – die Clusterkerne sind in Zeile 1 für alle Suchwortkombinationen gleich. In Normalschrift folgen jeweils in der Reihenfolge der Aufnahme in das Cluster die weiteren Seiten. Eine unterstrichene Wissensseinheit bedeutet, dass diese nicht mehr in das Cluster aufgenommen wurde, obwohl sie als ähnlichste verbleibende Wissensseinheit an der Reihe gewesen wäre. In diesem Fall wurde ein neues Cluster erzeugt, das in der jeweils unterhalb gelegenen Zelle dargestellt wird. Die Startseite dieses Clusters war jeweils die verbleibende freie Seite, die dem gerade abgeschlossenen Cluster am unähnlichsten war. Hierdurch soll erreicht werden, dass sich die Cluster untereinander möglichst unähnlich sind, um eine schärfere Themenabgrenzung zu erreichen.

Während die Startseite des ersten Clusters bei allen Kombinationen identisch war – und auch in drei der Fälle die zweite Seite dieses Clusters – war die erste Seite des zweiten Clusters bei allen Kombinationen eine andere. Dies lag sowohl am unterschiedlich aufgebauten ersten Cluster und dessen Schwerpunkt als auch an den noch verbleibenden Seiten – die nicht identisch waren – und deren Clusterdifferenzmaß zum ersten Cluster. Die Bildung von Clustern durch das Programm verlief unabhängig von den Suchbegriffen ähnlich – so ist das Cluster „Verarbeiten der Duroplaste-Polykondensationsprodukte“ in dreien der Kombinationen identisch.

Einige Schritte bzw. Punkte in den Clustering-Abläufen sollen besonders erläutert werden. So stellt der Umstand, dass es Cluster aus hierarchisch gleichberechtigten Seiten geben kann, eine Schwierigkeit bei der hierarchischen Gliederung der Seiten in einem Cluster dar, denn es muss immer eine Seite als Hauptseite ausgewählt werden. Dies sieht man am Punkt (1) für das Cluster aus den Seiten „Polymerisation – radikalisch“ und „Polymerisation – ionisch“. Beide Seiten bilden kor-

⁷⁴ Im Gegensatz zur Darstellung in Tabelle 6.8 wird hier auf die Angabe von Zahlenwerten verzichtet.

rekterweise aus sachlogischen Gründen einen Cluster. Sie sind hierarchisch prinzipiell gleichberechtigt. Trotzdem muss eine der beiden Seiten der anderen als Hauptseite übergeordnet werden. Eine sinnvolle Lösung für dieses Problem könnte eine Einbeziehung taxonomischer Kriterien in die Clusterbildung sein.

Tabelle 6.15: Clustering der verschiedenen Alternativen zu Lehrpfad 2

Nr.	Polykondensation – Polymerisation	Polyaddition – Polykondensation	Polyaddition – Polymerisation	Polyaddition – Polykondensation – Polymerisation
1	Einteilung nach dem Herstellungsprinzip Reaktion vom Monomeren zum Polymeren Kunststoffe <u>Einführung – Was sind Kunststoffe?</u>	Einteilung nach dem Herstellungsprinzip Reaktion vom Monomeren zum Polymeren <u>Einführung – Was sind Kunststoffe?</u>	Einteilung nach dem Herstellungsprinzip Reaktion vom Monomeren zum Polymeren Kunststoffe <u>Einführung – Was sind Kunststoffe?</u>	Einteilung nach dem Herstellungsprinzip Kurzzeichen für Kunststoffe (4) Reaktion vom Monomeren zum Polymeren Kunststoffe (4a)
2	Werkstoffe zur Verklebung mit Zweikomponentenklebern Einführung – Was sind Kunststoffe? <u>Verarbeiten der Duroplaste-Polykondensationsprodukte</u>	Polykondensation am Beispiel von Phenolharz (2a) Polyaddition Verarbeiten der Duroplaste-Polykondensationsprodukte (2) <u>Polykondensation</u>	Verarbeiten der Duroplaste-Polykondensationsprodukte (2) Polyaddition Polykondensation Vergleich: Polymerisation-Polykondensation-Polyaddition <u>Abbindevorgänge beim Kleben</u>	Zeitschiene der geschichtlichen Entwicklung von Kunststoffen Kunststoffe <u>Vergleich: Polymerisation-Polykondensation-Polyaddition</u>
3	Polymerisation – radikalisch (1) Polymerisation – ionisch <u>Polykondensation</u>	Werkstoffe zur Verklebung mit Zweikomponentenklebern Einkomponentenklebstoffe <u>Einführung – Was sind Kunststoffe?</u>	Polymerisation – radikalisch Polymerisation – ionisch <u>Einführung – Was sind Kunststoffe?</u>	Polymerisation – radikalisch Polymerisation – ionisch <u>Polykondensation</u>
4	Verarbeiten der Duroplaste-Polykondensationsprodukte (2) Polyaddition Polykondensation Vergleich: Polymerisation-Polykondensation-Polyaddition	Polykondensation (3) Vergleich: Polymerisation-Polykondensation-Polyaddition <u>Einführung – Was sind Kunststoffe?</u>	Abbinden mit chemischer Reaktion - Reaktionskleber und ihre Umgebungsbedingungen Abbindevorgänge beim Kleben <u>Einführung – Was sind Kunststoffe?</u>	Verarbeiten der Duroplaste-Polykondensationsprodukte Polyaddition Polykondensation Vergleich: Polymerisation-Polykondensation-Polyaddition <u>Abbinden mit chemischer Reaktion - Reaktionskleber und ihre Umgebungsbedingungen</u>
5		Einführung – Was sind Kunststoffe?	Einführung – Was sind Kunststoffe?	Abbinden mit chemischer Reaktion - Reaktionskleber und ihre Umgebungsbedingungen
fett: Clusterkerne, also Wissenseinheiten, die zuerst ein neues Cluster bilden <u>unterstrichen:</u> Wissenseinheiten, die in das betreffende Cluster nicht mehr aufgenommen werden (1-5): im Text diskutierte Punkte				

Eine weitere Besonderheit bei der Bestimmung der Reihenfolge der Cluster in einem Lehrpfad kann man an der Kombination „Polyaddition – Polykondensation“ erkennen. Dieser Lehrpfad ist wie auch „Polyaddition – Polymerisation“ und „Polykondensation – Polymerisation“ vom Abstrakten zum Konkreten sortiert – dies ist in Tabelle 6.14 dargestellt. In den letzteren beiden Lehrpfaden ist ein identisches Cluster mit der Hauptseite „Verarbeiten der Duroplaste-Polykondensationsprodukte“ (in Tabelle 6.15 mit (2) bezeichnet) enthalten. Dieses ist jeweils an zweiter Stelle der Reihenfolge positioniert. Im Lehrpfad „Polyaddition – Polykondensation“ wurde dieses Cluster aufgeteilt, da eine weitere Seite (2a) Bestandteil des ersten Teilclusters war. Eines der resultierenden Teilcluster steht nun in der Reihenfolge nach Tabelle 6.14 an erster und eines an letzter Stelle. Man sieht auf diese Weise anschaulich, dass sich die didaktischen Eigenschaften von Wissensseinheiten bei der Zusammenfassung gegenseitig aufheben können. Dies ist einer der Gründe, warum jeweils die aussagekräftigste der drei möglichen Anordnungen dem Lerner angezeigt wird.

Änderungen in der Zusammenstellung der im Lehrpfad enthaltenen Seiten können auch deren Clusterbildung beeinflussen. Wählt man ähnliche, aber andere Suchbegriffe, so werden der Mittelwert und die Streuung der Clusterähnlichkeitsmaße und damit der Grenzwert für die Clusterzugehörigkeit G_C beeinflusst. Dies kann zur Folge haben, dass Cluster in zwei Lehrpfaden unterschiedlich aufgebaut sind, obwohl die in dem Cluster des einen Lehrpfads enthaltenen Wissensseinheiten auch in dem zweiten Lehrpfad enthalten sind. Denselben Effekt wie eine Änderung der genannten Kennwerte hat das Vorhandensein einer „neuen“ Wissensseinheit, die wegen einer höheren Ähnlichkeit statt der „ursprünglichen“ Seite in ein Cluster aufgenommen wird. Ein Beispiel hierfür ist die Seite „Kurzzeichen für Kunststoffe“ (4), die den Schwerpunkt des Clusters änderte, sodass die Seite „Kunststoffe“ (4a) nicht mehr aufgenommen wurde. Analog kann eine „neue“ Seite auch Schwerpunkt eines Clusters werden und Teile von Seiten, die ansonsten ein Cluster gebildet hätten, an sich binden. Dies geschah z.B. bei der Seite „Polykondensation am Beispiel von Phenolharz“ (2a), was wie oben beschrieben zur Auftrennung des Clusters „Verarbeiten der Duroplaste-Polykondensationsprodukte“ (2) führte. Als die Wissensseinheit „Polykondensation“ aufgenommen werden sollte, war der Grenzwert schon zu klein, um die Seite noch in das Cluster aufzunehmen. Diese bildete dann ein eigenes Cluster (3).

Das Setzen der relevantesten Seite als ersten Clusterkern hat zur Folge, dass im ersten Cluster Wissensseinheiten gruppiert werden, die einem zentralen Thema ähnlich sind. Dies erhöht die Wahrscheinlichkeit, dass die für den Lehrpfad wichtigsten Themen in einem Cluster gesammelt werden.

Die Cluster müssen, selbst wenn die selben Seiten in einem Lehrpfad vorhanden sind, nicht immer identisch sein. Erschließt sich ein Lerner ein Themenfeld durch Generieren mehrerer Lehrpfade, so können dort durchaus Seiten enthalten sein, die in einem anderen Kontext in unterschiedlichen

Wissensclustern eingeordnet sind. Dies ist ein Sachverhalt, der dem Lerner mitgeteilt werden sollte, damit dieser mit dem System effizient lernen kann.

6.5 Lehrpfad 3

Die Thematik des dritten Lehrpfads war die Hydratation von Zement. Es wurde angenommen, dass ein fiktiver Lerner Informationen dazu benötigt, wie der Erhärtungsvorgang des Zementleims abläuft. Der erstellte Lehrpfad besteht aus 11 Seiten, die fünf Cluster bilden. Inhalte und Struktur des Lehrpfads können Tabelle 6.16 entnommen werden.

Tabelle 6.16: Zusammenstellung der untersuchten Lehrpfade

Suchbegriffe / Kenngrößen	Struktur des erzeugten Lehrpfads
<ul style="list-style-type: none"> ● Hydratation ● Zement <p>Sollwert L: 15 Seiten</p> <p>tatsächlicher Umfang: 11 Seiten</p> <p>Ordnungsprinzip: vom Allgemeinen zum Speziellen</p>	<ul style="list-style-type: none"> ► Zement <ul style="list-style-type: none"> - Rohstoffe - Nebenbestandteile des Zements ► Hydratation <ul style="list-style-type: none"> - Hydratationsstufe III ► Ansteifen, Erstarren & Erhärten ► Hydratationswärme <ul style="list-style-type: none"> - Schnelle Hydratationswärmeentwicklung - <i>abfließende Hydratationswärme</i> ► Hauptbestandteile des Zements <ul style="list-style-type: none"> - <i>Puzzolane</i> - <i>Begriffsbestimmung – Zement</i> - Hydratationsbeeinflussende Stoffe <ul style="list-style-type: none"> - <i>Zusammensetzung von Zement</i> - Hauptarten des Zements <ul style="list-style-type: none"> - <i>Vergleich alte und neue Norm</i> - Hauptbestandteile - Erhärten des Betons
Kennzeichnung der Strukturierung: ► Hauptseiten, - Nebenseiten, - <i>weiterführende Seiten</i>	

Der Lehrpfad „Hydratation – Zement“ ist inhaltlich und strukturell weitgehend vollständig und geeignet, eine Einführung in das Thema „Hydratation von Zement“ zu geben. Als offensichtlich fehlend fallen die Wissensseinheiten zu den Hydratationsstufen I und II auf. Diese Wissensseinheiten erreichten bei der Relevanzberechnung die Rangplätze 17 und 32. Sie fielen somit nicht unter die L relevantesten Seiten. Die sim_{UT} -Werte beider Seiten zum Unterthema Zement waren sehr niedrig, weshalb eine nachträgliche Aufnahme ebenfalls ausschied.

Mit diesem Lehrpfad wurde der Lehrpfad zu den Suchbegriffen „Abbinden“ – „Zement“ verglichen. Es kann davon ausgegangen werden, dass ein Lerner im Grundstudium bei beiden Suchanfragen denselben Lehrpfad erwarten würde. Dies ist aber objektiv nicht der Fall. Vielmehr thematisiert die

Variante des Lehrpfads neben dem Baustoff Zement auch Klebstoffe und deren Abbindeprozess. Mehrere der enthaltenen Seiten sind für das dem Lerner unterstellte Informationsbedürfnis irrelevant. Das unterstellte Lernziel des fiktiven Lernalters kann mit diesem Lehrpfad nicht erreicht werden. Der Aufbau beider Lehrpfade ist in Tabelle 6.17 gegenübergestellt.

Tabelle 6.17: Vergleich von Lehrpfad 3 und dem Lehrpfad zu den Suchbegriffen „Abbinden“ – „Zement“

„Hydratation“ – „Zement“	„Abbinden“ – „Zement“
<ul style="list-style-type: none"> ▶ Zement ▶ Hydratation* <ul style="list-style-type: none"> - Hydratationsstufe III* ▶ Ansteifen, Erstarren & Erhärten* ▶ Hydratationswärme* <ul style="list-style-type: none"> - Schnelle Hydratationswärmeentwicklung* ▶ Hauptbestandteile des Zements <ul style="list-style-type: none"> - Hydratationsbeeinflussende Stoffe* - Hauptarten des Zements - Hauptbestandteile - Erhärten des Betons* 	<ul style="list-style-type: none"> ▶ Verfahren der Zementherstellung <ul style="list-style-type: none"> - Das Verfahren der Zementherstellung ▶ Hauptbestandteile des Zements <ul style="list-style-type: none"> - Hauptbestandteile - Hauptarten des Zements ▶ Portlandzementklinker (K) <ul style="list-style-type: none"> - Zement ▶ Hydratationswärme* ▶ Systematik der Klebstoffe in Anlehnung an DIN 16 920 <ul style="list-style-type: none"> - Physikalisch abbindende Klebstoffe ▶ Abbindevorgänge beim Kleben ▶ Erhärten des Betons*
Sollwert L: 15 Seiten tatsächlicher Umfang: 11 Seiten Ordnungsprinzip: allgemein → speziell	Sollwert L: 13 Seiten tatsächlicher Umfang: 12 Seiten Ordnungsprinzip: abstrakt → konkret
Kennzeichnung der Strukturierung: ▶ Hauptseiten, - Nebenseiten *) Wissenseinheiten, die explizit die Hydratation des Zements betreffen	

Es soll nun untersucht werden, wie sich die beiden Lehrpfade zusammensetzen. Zunächst wird die Suchterweiterung betrachtet. Diese ist für beide Suchanfragen in Tabelle 6.18 dargestellt.

Tabelle 6.18: Suchterweiterungen für „Hydratation“ – „Zement“ und „Abbinden“ – „Zement“

Grundlage	Suchbegriffe	
	„Hydratation“ – „Zement“	„Abbinden“ – „Zement“
Original-Suchbegriffe	hydratation, zement	abbinde, zement
paarweise gemeinsame ähnliche Begriffe	treiberscheinungen, zementzusammensetzung, erhaftungszustand, puzzolanement, bricht, treib, exotherm, ansteifen, raumbestaendigkeit, anmachwasser, storer, waermeentwicklung, erhaert, fehlerhaft, hervorgerufen, wasserzementwert, zementstein, erstarr, hydrationswaermeentwicklung, zementleim, mahlfein, hydratation, hydrationswaerme, zementkoerner, leim, erstarrungsstoerungen	–
gemeinsame Kookkurrenten	land, hydr, gebundem, erstarrungsstoerungen, beruhend, puzzolanement, calcium, temperaturentwicklung, erhaertungsreakti-	beruhend, strich, massig, bestand, anhydritestrich, anmachwasser, leim, zementestrich, bestandteil, hydratation, umgeb,

Grundlage	Suchbegriffe	
	„Hydratation“ – „Zement“	„Abbinden“ – „Zement“
	on, lehrgang, portland, weswegen, zementkoerner, hydrationswaermeentwicklung, jungen, zementzusammensetzung, hydroxid, treiberscheinungen, portlandoelschieferzement, hs, zementfestigkeitsklasse, klinkerminerale, klinker, hochofenzement, hydrationsprodukt, tragwerksplan, stoer, waermeentwicklung, nacherhaertung, hydratphasen, massig, hydrationsstuf, anmachwasser, portlandzement, treib, vorteilhaft, zusatzmittel, calciumhydroxid, leim, puzzolan, erschein, vermisch, festigkeitsentwicklung, zeitraum, freigesetzt, wasserzementwert, iii, por, hydrationswaerme, festigkeitsklasse, gebunden, setz, bezogen, zementleim, zementstein, kontakt, alt, frei, erhaert, tag, zusammensetzung, cem, erstarr, entwickl, lehrpfad, schnell, infolge, waehrend, entsteh, prozent, und	estrich, kontakt, erhaert, erstarr, infolge, erforderlich, waehrend
höchstsignifikante Kookkurrenten der Original-Suchbegriffe	hydrationswaerme, zement, hydrationsgrad, hydrationsstuf, waermeentwicklung, stuf, warm, hydrationswaermeentwicklung, entwickl, hydrationsprodukt, zementleim, zementstein, leim, portlandzement, portland, wasser, klinker, zementklinker, por, land	abbindevorgang, klebstoff, kleb, vorgang, reaktion, polyacrylate, chemisch, physikalisch, reaktionsklebstoff, gummi, zementleim, zementstein, leim, portlandzement, portland, wasser, klinker, zementklinker, por, land

Der erweiterte Suchvektor zur Kombination „Hydratation – Zement“ hat einen Umfang von 85 Einträgen, der zu „Abbinden – Zement“ einen Umfang von 38 Einträgen. Dabei sind die Einträge des Suchvektors zu „Hydratation – Zement“ näher auf das implizierte Thema bezogen als die zu „Abbinden – Zement“. Gefunden wurden zu „Hydratation – Zement“ 1327 Wissenseinheiten, zu „Abbinden – Zement“ 921 Wissenseinheiten.

Im fertigen Lehrpfad „Hydratation – Zement“ finden sich sechs Wissenseinheiten, die explizit das Thema „Hydratation von Zement“ zum Inhalt haben, im Lehrpfad „Abbinden – Zement“ finden sich zwei. Es interessiert also die Frage, wieso diese Wissenseinheiten im Lehrpfad zu „Hydratation – Zement“ vorhanden sind, in dem zu „Abbinden – Zement“ jedoch nicht. Außerdem ist zu klären, wie die Wissenseinheiten zum Thema Klebstoff in den Lehrpfad zu „Abbinden – Zement“ gelangten. Dabei ist zu beachten, welche Wissenseinheiten jeweils nachträglich entfernt und eingefügt wurden. Betrachtet werden exemplarisch die Wissenseinheiten „Systematik der Klebstoffe in Anlehnung an DIN 16920“, „Hydrationswärme“, „Hydratation“ und „Physikalisch abbindende Klebstoffe“. Deren Kennwerte für die Relevanzberechnung sind in Tabelle 6.19 dargestellt.

Die Wissensseinheit „Systematik der Klebstoffe in Anlehnung an DIN 16920“ weist eine hohe Ähnlichkeit sim_{OV} mit den ursprünglichen Suchbegriffen auf. Bei der Ähnlichkeit sim_{SV} zum Suchvektor ist ihre Zugehörigkeit zum mittleren Bereich am größten. Die hohe Ähnlichkeit mit dem Kontext der auf der Seite vorhandenen Suchbegriffe bewirkt eine große Zugehörigkeit zum hohen Bereich. Werte durchaus vergleichbarer Größenordnung treten auch bei der Wissensseinheit „Hydratationswärme“ auf. Beachtenswert sind auch die Werte für sim_{UT} : einmal ist der Wert für Zement („z“) sehr hoch, einmal der für „Abbinden“ („a“).

Tabelle 6.19: Kennwerte der Relevanzberechnung zur Suchanfrage „Abbinden – Zement“

Rang / Relevanz R	Titel	$\text{sim}_{\text{SV}}(\text{d}_i) / \mu_{\text{SV}}$	$\text{sim}_{\text{FV}}(\text{d}_i) / \mu_{\text{FV}}$	$\text{sim}_{\text{OV}}(\text{d}_i) / \mu_{\text{OV}}$	$A_{\text{FT}} / \mu_{\text{FT}}$	$A_{\text{UE}} / \mu_{\text{UE}}$	$\text{sim}_{\text{UT},n}(\text{d}_i)$
7. 69,31	Systematik der Klebstoffe in Anlehnung an DIN 16920	0,209 $\mu_{\text{SV,gering}} = 0,0$ $\mu_{\text{SV,mittel}} = 0,88$ $\mu_{\text{SV,hoch}} = 0,12$	0,7 $\mu_{\text{FV,gering}} = 0,0$ $\mu_{\text{FV,mittel}} = 0,04$ $\mu_{\text{FV,hoch}} = 0,96$	0,35 $\mu_{\text{OV,gering}} = 0,0$ $\mu_{\text{OV,mittel}} = 0,08$ $\mu_{\text{OV,hoch}} = 0,92$	50 % $\mu_{\text{FT,gering}} = 0,0$ $\mu_{\text{FT,mittel}} = 0,5$ $\mu_{\text{FT,hoch}} = 0,5$	0 % $\mu_{\text{UE,gering}} = 1,0$ $\mu_{\text{UE,mittel}} = 0,0$ $\mu_{\text{UE,hoch}} = 0,0$	a: 0,325 z: 0,0
9. 66,57	Hydratationswärme	0,244 $\mu_{\text{SV,gering}} = 0,0$ $\mu_{\text{SV,mittel}} = 0,63$ $\mu_{\text{SV,hoch}} = 0,37$	0,52 $\mu_{\text{FV,gering}} = 0,0$ $\mu_{\text{FV,mittel}} = 0,74$ $\mu_{\text{FV,hoch}} = 0,26$	0,306 $\mu_{\text{OV,gering}} = 0,0$ $\mu_{\text{OV,mittel}} = 0,35$ $\mu_{\text{OV,hoch}} = 0,65$	100 % $\mu_{\text{FT,gering}} = 0,0$ $\mu_{\text{FT,mittel}} = 0,0$ $\mu_{\text{FT,hoch}} = 1,0$	0 % $\mu_{\text{UE,gering}} = 1,0$ $\mu_{\text{UE,mittel}} = 0,0$ $\mu_{\text{UE,hoch}} = 0,0$	a: 0,02 z: 0,192
16.* 59,61	Hydratation	0,253 $\mu_{\text{SV,gering}} = 0,0$ $\mu_{\text{SV,mittel}} = 0,57$ $\mu_{\text{SV,hoch}} = 0,43$	0,45 $\mu_{\text{FV,gering}} = 0,02$ $\mu_{\text{FV,mittel}} = 0,92$ $\mu_{\text{FV,hoch}} = 0,0$	0,347 $\mu_{\text{OV,gering}} = 0,0$ $\mu_{\text{OV,mittel}} = 0,1$ $\mu_{\text{OV,hoch}} = 0,9$	100 % $\mu_{\text{FT,gering}} = 0,0$ $\mu_{\text{FT,mittel}} = 0,0$ $\mu_{\text{FT,hoch}} = 1,0$	0 % $\mu_{\text{UE,gering}} = 1,0$ $\mu_{\text{UE,mittel}} = 0,0$ $\mu_{\text{UE,hoch}} = 0,0$	a: 0,206 z: 0,12
19.** 57,94	Physikalisch abbindende Klebstoffe	0,148 $\mu_{\text{SV,gering}} = 0,32$ $\mu_{\text{SV,mittel}} = 0,68$ $\mu_{\text{SV,hoch}} = 0,0$	0,583 $\mu_{\text{FV,gering}} = 0,0$ $\mu_{\text{FV,mittel}} = 0,49$ $\mu_{\text{FV,hoch}} = 0,51$	0,23 $\mu_{\text{OV,gering}} = 0,0$ $\mu_{\text{OV,mittel}} = 0,81$ $\mu_{\text{OV,hoch}} = 0,19$	50 % $\mu_{\text{FT,gering}} = 0,0$ $\mu_{\text{FT,mittel}} = 0,5$ $\mu_{\text{FT,hoch}} = 0,5$	50 % $\mu_{\text{UE,gering}} = 0,0$ $\mu_{\text{UE,mittel}} = 0,5$ $\mu_{\text{UE,hoch}} = 0,5$	a: 0,24 z: 0,01
*) relevante Wissensseinheit, die nicht in den Lehrpfad aufgenommen wird							
**) Wissensseinheit, die nachträglich in den Lehrpfad aufgenommen wird							

Eine zweifelsfrei relevante Seite für den Lehrpfad „Abbinden – Zement“ ist die Seite „Hydratation“. Diese wurde jedoch nicht in den Lehrpfad aufgenommen. Während sowohl ihr sim_{SV} -Wert als auch ihr sim_{OV} -Wert höher sind als bei der zuvor diskutierten Wissensseinheit, ist einzig der sim_{FV} -Wert auf einem niedrigeren Level angesiedelt und bestimmt dadurch den niedrigeren Relevanzwert. Beide sim_{UT} -Werte sind hierbei nicht hoch genug für eine nachträgliche Aufnahme der Wissensseinheit in den Lehrpfad. Die Seite „Physikalisch abbindende Klebstoffe“ weist durchweg niedrige Eingangswerte auf. Sie wurde jedoch durch die hohe Ähnlichkeit mit dem Unterthema „Abbinden“ nachträglich in den Lehrpfad aufgenommen.

Die Seite „Zement in der Neuzeit - moderne Zemente“ wurde vom Programm aus beiden Lehrpfaden nachträglich entfernt, da sowohl ihr sim_{UT} -Wert zum Suchbegriff „Abbinden“ als auch der zum Suchbegriff „Hydratation“ sehr gering sind. Aus dem Lehrpfad „Hydratation – Zement“ wurde ferner

die Seite „Portlandzementklinker“ entfernt, ebenfalls wegen eines zu geringen sim_{UT} -Werts zum Suchbegriff „Hydratation“.

Schon bei den dargestellten vier Wissensseinheiten fällt auf, dass bei den Ähnlichkeiten zu den Unterthemen häufig der Wert null auftritt. Bei den L relevantesten Wissensseinheiten zum Lehrpfad „Abbinden – Zement“ kommt dies acht mal vor, bei „Hydratation – Zement“ jedoch gar nicht. Prinzipiell kann der Lehrpfad „Abbinden – Zement“ auch als Lehrpfad zu den beiden getrennten Themen „Abbinden“ und „Zement“ betrachtet werden.

Wegen der häufigen Werte für $\text{sim}_{\text{UT}} = 0$ kann man vermuten, dass die Erstellung des Lehrpfads viele irrelevante Wissensseinheiten zum implizierten, übergreifenden Thema erbracht hat, weil die Suchbegriffe eine zu geringe Verbindung untereinander aufweisen. Es handelt sich bei „Hydratation“ und „Abbinden“ um eine Hyperonym-Hyponym-Relation, d.h. Hydratation ist ein Unterbegriff von Abbinden. Abbinden umfasst deshalb ein breiteres Themenfeld als Zement, u.a. – wie in diesem Beispiel – Klebstoffe. Die Begriffe Abbinden und Zement sind also nicht so eng verzahnt wie die Begriffe „Hydratation“ und Zement. Das Ergebnis ist ein Lehrpfad, der durch einen hohen Anteil nicht relevanter Seiten charakterisiert ist.

Diese Vermutung wird im nächsten Kapitel detailliert untersucht.

7 Zusammenhang zwischen Suchanfrage und Lehrpfadqualität

7.1 Notwendigkeit und Möglichkeiten der Bewertung

Die Implementierung des entwickelten Vorschlags zur automatischen Generierung von Lehrpfaden liefert Lerneinheiten, die in der Praxis durch Studierende und Lehrende genutzt werden können. Die Lehrpfade können sich jedoch in ihrer inhaltlichen und didaktischen Qualität unterscheiden. Dies kann – die inhaltliche Korrektheit der einzelnen Wissenseinheiten vorausgesetzt – verschiedene Ursachen haben:

- **Umfang der Wissenseinheiten:** Für die Antwort auf die Suchanfrage eines Nutzers können zu wenige Inhalte vorhanden sein. Die wenigen relevanten Inhalte können im Lehrpfad außerdem mit irrelevanten Inhalten ergänzt sein. Auch ein Überangebot an relevanten Inhalten zu einer Suchanfrage kann die Antwort verschlechtern, falls unter der Vielzahl der Wissenseinheiten die relevantesten nicht identifiziert werden können.
- **Formulierung der Suchanfrage:** Ein Lerner kann seine Suchanfrage falsch formuliert haben, sodass das System eine Antwort gibt, welche eigentlich eine andere Frage betrifft, als die, die der Lerner beabsichtigt hatte.
- **Verknüpfung der Begriffe:** Begriffe können fehlerhaft miteinander verknüpft sein. Bei sehr kurzen Sätzen in den Wissenseinheiten können jeweils nur wenige Begriffe als lokaler Kontext verarbeitet werden. Bei sehr langen Sätzen erhöht sich das Signifikanzmaß der Konkurrenz zwischen miteinander nicht verwandten Begriffen. Die Verwendung unterschiedlicher Fachbegriffe für denselben Sachverhalt verschlechtert ebenfalls die Verknüpfungen zwischen Begriffen.
- **Sonderfälle:** Bei speziellen Begriffen, die in mehreren Kontexten vorkommen, können irrelevante Wissenseinheiten auftreten, die keinen Bezug zu den anderen Suchbegriffen haben.

Lerner, insbesondere solche mit geringen Vorkenntnissen, benötigen qualitativ möglichst hochwertige Lerninhalte. Es stellt sich also die Frage nach einer Beurteilung der Qualität der erstellten Lehrpfade. Leider entziehen sich einzelne generierte Lehrpfade einer objektiven Bewertung. Sie entfalten ihren didaktischen Mehrwert nur subjektiv für den Nutzer und sind damit u.a. abhängig von dessen Motivation, Vorkenntnissen und Lernverhalten. Aufgrund dieser Diversität der Studierenden hängt der Lernerfolg für ein bestimmtes Lernelement stark vom jeweiligen Studierenden ab, das Lernerlebnis ist ebenfalls subjektiv. Gefragt sind also objektive Beurteilungskriterien für die Qualität der generierten Lehrpfade. Folgende Kriterien stehen dafür grundsätzlich zur Verfügung:

- Ein Vergleich der automatisch erzeugten Lehrpfade mit entsprechenden Lehrpfaden des WiBA-Net, die von Autoren erstellt wurden. Hier ergibt sich aber das Problem, dass das Vermittlungsziel der manuell erstellten Lehrpfade nicht mit dem durch eine Suchanfrage implizierten Lernziel identisch sein muss.
- Die Durchführung eines Relevance Feedback [FERBER 2003]. Jeder Nutzer kann zu den erstellten Lehrpfaden seine Meinung abgeben, bzw. die Relevanz des Lehrpfads für sein Problem beurteilen. Dieses Verfahren bringt zwei Probleme mit sich. Zum einen kann ein Nutzer mit geringem Vorwissen nicht hinreichend beurteilen, ob ihm die für ein konkretes Problem relevanten Inhalte vermittelt wurden. Zum anderen müsste letztlich die Relevanz jeder einzelnen im Lehrpfad enthaltenen Wissensseinheit abgefragt werden, damit ein etwaiger Fehler eingegrenzt werden kann.
- Die manuelle Auswertung einer Reihe von automatisch generierten Lehrpfaden. Für jede Wissensseinheit wird notiert, ob sie für das Thema relevant ist oder nicht. Die Qualität der Lehrpfade kann dann anhand des Anteils der relevanten Wissensseinheiten – der Precision – beurteilt und vergleichbar gemacht werden.

Das letzte genannte Verfahren ist zwar nicht objektiv, bietet jedoch die Möglichkeit, durch eine entsprechende Dokumentation und Diskussion die Ergebnisse der Auswertung transparent und nachvollziehbar zu machen. Aus diesem Grund wird zur Bewertung der inhaltlichen Qualität der automatisch erstellten Lehrpfade eine manuelle Auswertung verschiedener Lehrpfade gewählt.

Die Analyse der Ergebnisse der Lehrpfaderstellung in Kapitel 6 ließ die Vermutung aufkommen, dass die inhaltliche Qualität eines Lehrpfads von den verwendeten Suchbegriffen abhängen könnte. Unter diesem Gesichtspunkt sollen Rückschlüsse auf die Art der Eingabeparameter gezogen werden, die zu qualitativ guten Lehrpfaden führen. Diese könnten dann vom Anwender beeinflusst werden, bzw. es kann der Anwender vor dem Arbeiten mit dem Lehrpfad darüber informiert werden, wie relevant die Inhalte des Lehrpfads für seine Suchanfrage wahrscheinlich sind.

Im Folgenden wird eine statistische Bewertung mehrerer automatisch erzeugter Lehrpfade hinsichtlich ihrer inhaltlichen Relevanz, die auf der manuellen Auswertung der Lehrpfade beruht, durchgeführt. Diese Relevanz als Kriterium für die Qualität der Lehrpfade kann mit bestimmten Eigenschaften der jeweiligen Suchbegriffe korreliert werden. Daraus lässt sich ein Vorhersagekriterium für die Relevanz von Lehrpfaden ableiten.

Eine Information über die zu erwartende Relevanz eines Lehrpfads als Feedback des Systems bei einer Suchanfrage ist insbesondere bei Lernern mit keinen oder nur geringen fachlichen Vorkenntnissen wichtig. Diese können die Relevanz der in einem erstellten Lehrpfad enthaltenen Inhalte nicht selbst beurteilen. Dies gilt mindestens, bis sie sie durch ausreichenden Kenntniserwerb in

einen Kontext mit anderen Inhalten einordnen können. Wenn Lerner nicht sicher sein können, die richtigen Informationen zu erhalten, weil sie nicht wissen, welche sie überhaupt benötigen, ist dies ein Umstand, der die Lernmotivation senken kann. Für Lehrende, die das System nutzen, ist weniger die Einschätzung der Relevanz der Inhalte wichtig, als ein Anhaltspunkt, durch welche Art der Eingabe ein möglichst hoch-relevanter Lehrpfad erzeugt werden kann, um den für die Nachbearbeitung erforderlichen Aufwand zu reduzieren.

Die Information, wie hoch der Anteil an relevanten Wissensseinheiten an einem Lehrpfad ist, kann für die Lerner ergänzt werden. So können Wissensseinheiten, die nur einen geringen Relevanzwert aufweisen (z.B. unter einem bestimmten Grenzwert) oder deren Relevanzwert eine maximale Differenz zum nächsthöheren Relevanzwert überschreitet, im Lehrpfad als potentiell nicht relevante Seiten farblich markiert werden.

7.2 Relevanz der Lehrpfade bezüglich Suchanfragen

Die vom System generierten Lehrpfade unterscheiden sich qualitativ. Dies kann wie in Abschnitt 7.1 beschrieben verschiedene Ursachen haben und sich entweder in den Inhalten oder in der Zusammenstellung der Wissensseinheiten manifestieren. Die auf didaktischen Kriterien beruhende Zusammenstellung der Seiten soll an dieser Stelle nicht weiter betrachtet werden, da die sich darin ausdrückende Qualität eines Lehrpfades nicht eindeutig beurteilt und quantifiziert werden kann. Damit ein Lerner die didaktische Struktur des Lehrpfades selbst einordnen und seinen Lernvorgang an ihr ausrichten kann, wird ihm als Hilfestellung das dem Lehrpfad zugrunde liegende Ordnungsprinzip mitgeteilt.

Die inhaltliche Qualität kann ebenfalls nicht allgemein verbindlich beurteilt werden. Für sie ist aber zumindest eine subjektive Beurteilung möglich, die sich außerdem in konkreten Zahlenwerten ausdrücken lässt und damit nachvollziehbar wird. Auf der Basis dieser Beurteilung kann dem Nutzer ebenfalls eine Hilfestellung zur Einschätzung der Qualität eines Lehrpfades gegeben werden. Die Bestimmung und Bewertung der inhaltlichen Qualität von Lehrpfaden wird im Folgenden erläutert.

Die Analyse verschiedener Lehrpfade legt die Vermutung nahe, dass die Qualität eines Lehrpfades von der Art der Suchanfrage abhängig ist. Konkret wird angenommen, dass das System bessere Lehrpfade produziert, je enger die einzelnen Suchbegriffe einer Suchanfrage untereinander verknüpft sind. Im Folgenden soll diese Vermutung überprüft werden und bei einem positiven Ergebnis ein Kriterium zur Vorhersage der Qualität eines Lehrpfades auf Basis der Suchbegriffe abgeleitet werden. Diese prognostizierte Qualität soll dann den Anwendern, insbesondere den Lernern, bei der automatischen Erstellung eines Lehrpfades mitgeteilt werden, damit diese ihr Lernverhalten darauf einstellen können. Die errechnete Qualität wird dabei durch einen Zahlenwert ausgedrückt,

der auf einem bestimmten Signifikanzniveau angibt, wie hoch die aufgrund der Eingabe des Nutzers zu erwartende Relevanz der erstellten Lehrpfadinhalte ist.

Um eine Abhängigkeit zwischen Relevanz und Verknüpfung der Suchbegriffe darstellen zu können, muss zunächst eine Methode der Bewertung der Relevanz festgelegt und eine hinreichende Anzahl von Lehrpfaden diesbezüglich analysiert werden. Ferner muss die Verknüpfungsstärke zwischen einzelnen Suchbegriffen quantifiziert werden.

Für die Relevanz der zu einem Lehrpfad zusammengestellten Seiten kann keine wirklich objektive Beurteilung abgegeben werden, da diese für jede Suchanfrage neu für jede Seite zu bestimmen wäre und von der Einschätzung des jeweils beurteilenden Experten abhängt [FERBER 2003]. Eine automatische Qualitätsbestimmung anhand der berechneten Relevanz – z.B. mittels des durchschnittlichen berechneten Relevanzwerts – ist nicht geeignet. Dies liegt daran, dass auch Wissenseinheiten mit durchschnittlich niedrigen Relevanzwerten akzeptable Lehrpfade ergeben können (vgl. Anhang J). Außerdem ist die Größe der Relevanzwerte ein Teil des Algorithmus, dessen Qualität letztlich an dieser Stelle bewertet werden soll.

Aus diesem Grund wird eine subjektive Betrachtung der Relevanz der Wissenseinheiten verschiedener Lehrpfade vorgenommen. Zu verschiedenen Suchbegriffskombinationen werden Lehrpfade erstellt. Die „Bezugsgröße“ der Relevanzbeurteilung ist das vermutete Lernziel eines fiktiven Lerner, der die Suchanfrage eingibt. Von diesem Lernziel wird angenommen, dass es die Gesamthematik der eingegebenen Suchbegriffe in ihrem Kontext umfasst – die einzelnen Suchbegriffe sollten also nur einführend oder auf die Gesamthematik bezogen behandelt werden, nicht vertiefend, ohne dass sie einen Bezug zu den anderen Suchbegriffen haben.

Die Gliederung und die Reihenfolge der Wissenseinheiten als didaktische Kriterien werden an dieser Stelle nicht berücksichtigt, es wird ausschließlich die Zugehörigkeit einer Wissenseinheit zum Lehrpfad betrachtet. Da die Reihenfolge keine Rolle spielt, werden die Wissenseinheiten zunächst nach ihrem jeweiligen berechneten Relevanzwert geordnet. Es wird dann für jede Wissenseinheit angegeben, ob sie auf der Grundlage ihres Inhalts für den Lehrpfad nach der obigen Definition relevant ist oder nicht (vgl. Anhang J). Daraus kann für die einzelnen Lehrpfade ihr jeweiliger prozentualer Anteil an relevanten Seiten – ihre Precision P – berechnet werden.

Es werden ausschließlich die im Lehrpfad vorhandenen Wissenseinheiten betrachtet. Eventuell zur Vollständigkeit fehlende Inhalte werden nicht berücksichtigt. Der Recall-Wert des jeweiligen Lehrpfads wird also nicht ermittelt. Die Beschränkung auf die im Lehrpfad vorhandenen Seiten macht die Nachvollziehbarkeit der Relevanzbeurteilung einfacher. Es ist jedoch zu beachten, dass die „wahre“ Qualität eines Lehrpfads schlechter sein kann als angegeben, da Seiten zur vollständigen Abdeckung des Lernziels fehlen können.

Bei der Beurteilung der Relevanz wird grundsätzlich von einer scharfen Trennung zwischen relevant und irrelevant ausgegangen. Relevant sind ausschließlich Wissensseinheiten, die dem oben definierten Lernziel inhaltlich direkt entsprechen. Wissensseinheiten die diesem nicht direkt entsprechen, jedoch durchaus als weiterführende Seiten infrage kämen, weil sie die Thematik des Lehrpfads sinnvoll ergänzen, werden genauso als irrelevant betrachtet wie Wissensseinheiten, die überhaupt nichts mit dem Thema des Lehrpfads zu tun haben⁷⁵. Demgegenüber ist es für die Relevanz einer Wissensseinheit unerheblich, ob bereits eine Wissensseinheit mit der gleichen Aussage Bestandteil des Lehrpfads ist.

Die genaue Relevanzbeurteilung aller untersuchten Lehrpfade findet sich in Anhang J. In Tabelle 7.1 sind die errechneten Precision-Werte P für die einzelnen Suchbegriffskombinationen zusammengestellt.

Auf der anderen Seite müssen die maßgeblichen textstatistischen Eigenschaften von Suchbegriffen betrachtet werden, um ein Maß für die Verknüpfungsstärke innerhalb einer Suchwortkombination definieren zu können. Diese Eigenschaften sind der Signifikanzwert zwischen jeweils zwei Begriffen sowie die Häufigkeit der einzelnen Begriffe und die Umfänge von deren globalen Kontexten. Um aus diesen Größen einen gemeinsamen Kennwert zu berechnen, wird zunächst der Zusammenhang der Einflussgrößen betrachtet. Abbildung 7.1 stellt den durchschnittlichen Signifikanzwert zwischen einem Begriff A und allen seiner Kookkurrenten in Abhängigkeit von der Anzahl dieser Kookkurrenten dar. Die Anzahl der Kookkurrenten eines Begriffs entspricht dabei dem Umfang seines globalen Kontextes. Man erkennt, dass Signifikanzwert und Kookkurrentenanzahl gegenläufig orientiert sind: Mit steigendem Umfang des globalen Kontextes eines Begriffs sinkt dessen durchschnittliche Signifikanz zu dessen Kookkurrenten.

In Abbildung 7.2 wird der Signifikanzwert zweier Begriffe A und B in Abhängigkeit von der Größe $x = ab/n$ dargestellt (vgl. Formel (2.6)). Für eine größere Anzahl Sätze a oder b , in denen die Begriffe auftreten, wird der Signifikanzwert kleiner. Der Scharparameter ist die Häufigkeit k des gemeinsamen Auftretens der beiden Begriffe in einem Satz. Die einzelnen Graphen haben unterschiedliche Anfangspunkte, weil der jeweilige k -Wert des gemeinsamen Auftretens erst ab einer Mindestgröße von a und b auftreten kann. Man erkennt deutlich den Anstieg der Signifikanzwerte mit größer werdendem k -Wert. Das Diagramm zeigt, dass die Signifikanz und die Häufigkeit (charakterisiert durch die Anzahl der Sätze) der Begriffe ebenfalls gegenläufig orientiert sind.

Es wird davon ausgegangen, dass eine große Häufigkeit der Suchbegriffe eine niedrigere Relevanz des Lehrpfads bewirkt. Eine Suchanfrage mit sehr häufigen Suchbegriffen muss eine größere Auswahl an Wissensseinheiten berücksichtigen als eine Anfrage zu seltenen Themen. Ferner wird

⁷⁵ Auf die der Thematik ähnlichen Seiten wird allerdings im folgenden Abschnitt noch einmal gesondert eingegangen.

angenommen, dass ein Lehrpfad unpräziser wird, wenn der globale Kontext der Suchbegriffe groß wird. In diesem Fall werden durch die dann umfangreichere Suchtermerweiterung mehr Wissenseinheiten gefunden, die hinsichtlich ihrer Relevanz beurteilt werden müssen. Ein Begriff mit durchschnittlicher Häufigkeit, aber einem sehr großen globalen Kontext könnte mehrere Bedeutungen mit unterschiedlichen Kontexten aufweisen. Eine solche Kombination würde ebenfalls zu einer niedrigen Relevanz des Lehrpfades führen.

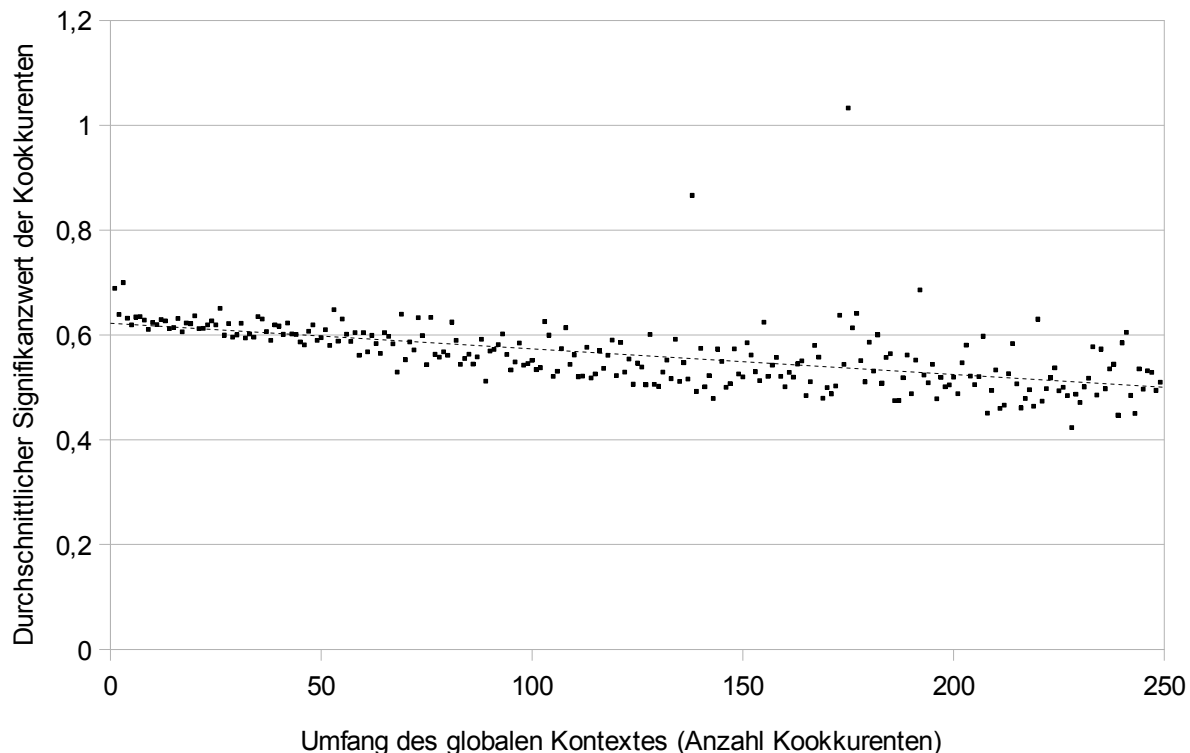


Abbildung 7.1: Durchschnittlicher Signifikanzwert der Kookkurrenten in Abhängigkeit von der Anzahl der Kookkurrenten zu einem Begriff (Regressionsgerade dient der Veranschaulichung)

Der Signifikanzwert zweier Begriffe bedeutet ein häufiges gemeinsames Auftreten der beiden Begriffe im selben Kontext. Je höher dieser Signifikanzwert ist, desto leichter können Querschnittsthemen erschlossen werden und desto höher fällt die Relevanz eines Lehrpfades aus. Der Signifikanzwert hat zwar eine hohe Aussagekraft, ist jedoch von der Häufigkeit der beiden Begriffe abhängig. Er eignet sich also nicht als alleinige Vergleichsgröße.

In die weitere Berechnung gehen die mittlere Kookkurrentenanzahl K^{76} und die mittlere Häufigkeit h aller Suchbegriffe einer Suchanfrage ein. Der Signifikanzparameter S wird durch Mittelung der Si-

76 Für die Darstellung in Abbildung 7.1 wurden nur Begriffe mit maximal 249 Kookkurrenten ausgewertet, da die maximale Anzahl von Kookkurrenten bei der Relevanzberechnung ebenfalls auf 250 beschränkt ist. Für die Bestimmung der mittleren Kookkurrentenanzahl K werden jedoch alle Kookkurrenten der jeweiligen Begriffe herangezogen.

gnifikanzwerte $\text{sig}(A,B)$ aller möglichen Kombinationen aus jeweils zwei Suchbegriffen der Suchanfrage gebildet.

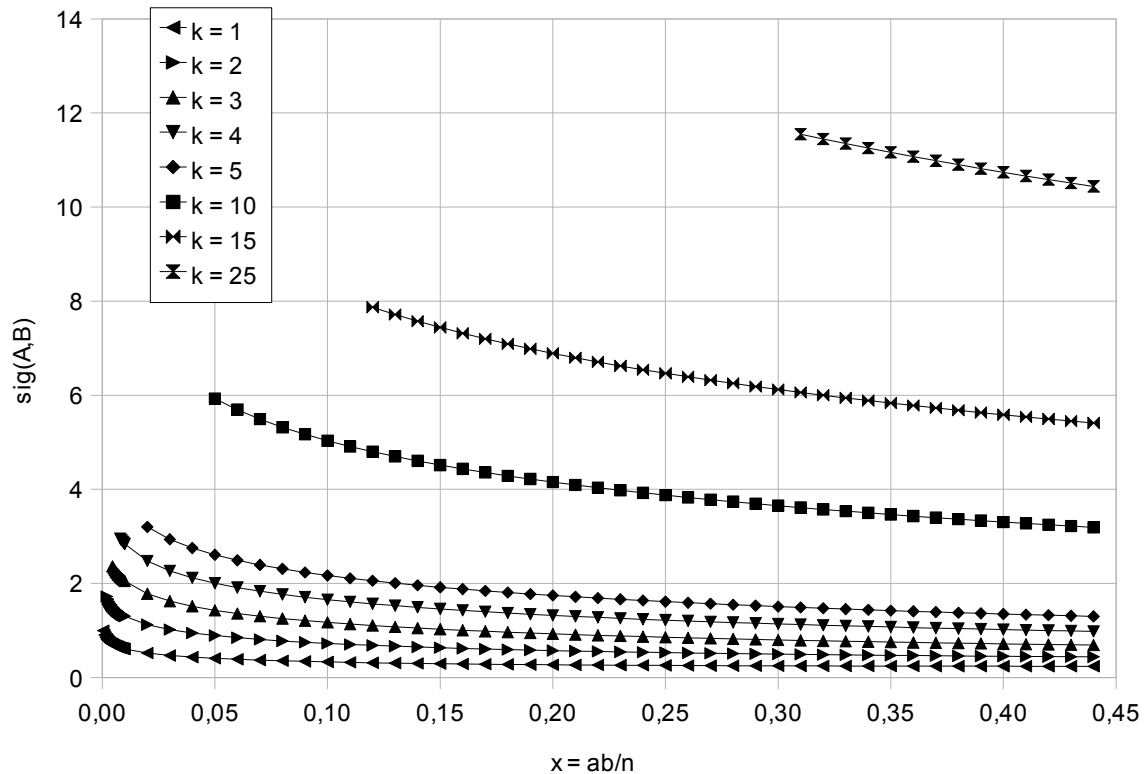


Abbildung 7.2: Signifikanz zweier Begriffe A und B in Abhängigkeit von den Anzahlen a und b der Sätze, in denen sie auftreten (Scharparameter k: Anzahl der Sätze, in denen A und B gemeinsam auftreten)

Mit diesen Größen kann jetzt ein Maß für die kontextbezogene Verknüpfungsstärke VS der Suchbegriffe untereinander definiert werden. Es wird vermutet, dass mit größerer Verknüpfungsstärke VS auch die Relevanz des erzeugten Lehrpfads ansteigt. Es bietet sich also eine Darstellung der Verknüpfungsstärke als Quotient aus S und dem Produkt aus K und h an. Für die Verknüpfungsstärke VS wird definiert⁷⁷:

$$VS = \frac{S \cdot 10.000}{K \cdot h} \quad (7.1)$$

Die Verknüpfungsstärke VS der Suchbegriffe und die Precision P der untersuchten Lehrpfade sind in Tabelle 7.1 verzeichnet. Dort sind ferner die mittleren Signifikanzwerte S, die mittleren Kookkurrenzenanzahlen K und die mittleren Häufigkeiten h aufgeführt. Die letzte Spalte beinhaltet den Wert P', der in Abschnitt 7.3 diskutiert wird.

⁷⁷ Der Faktor 10.000 hat keinen Einfluss auf den Korrelationskoeffizienten und dient lediglich der besseren Darstellbarkeit der Ergebnisse.

Tabelle 7.1: Übersicht der untersuchten Suchbegriffskombinationen mit ihren Kenngrößen

Nr.	Suchbegriffe	S	K	h	VS	P	P'
1	Beton – Brücke	0,1040	3194,50	363,50	0,0009	15,38%	46,15%
2	Definition – Dichte	0,0270	1463,00	102,50	0,0018	28,57%	42,86%
3	Beton – Ausgangsstoff	0,3130	2628,00	350,00	0,0034	68,75%	75,00%
4	Bruch – Stahlbeton	0,0260	1201,00	57,50	0,0038	42,86%	64,29%
5	Festigkeit – Legierung – Stahl	1,1260	2993,00	382,67	0,0098	54,55%	77,27%
6	Erzeugung – Stahl	1,2175	2653,50	390,50	0,0117	23,53%	47,06%
7	Herstellung – Hochofen – Stahl	0,9645	2381,33	329,00	0,0123	80,00%	80,00%
8	Beton – Bruch – Stahl	2,3377	3630,00	475,33	0,0135	58,82%	70,59%
9	Herstellung – Stahl	2,2515	3332,50	477,00	0,0142	60,00%	66,67%
10	Herstellung – Zement	0,8125	2206,50	259,50	0,0142	64,29%	71,43%
11	Erzeugung – Hochofen – Stahl	0,7502	1928,67	271,33	0,0143	80,00%	80,00%
12	Holz – Verbindung	2,0620	3410,00	334,50	0,0181	63,16%	68,42%
13	Beton – Korrosion – Stahl	3,4110	3806,33	491,67	0,0182	63,16%	68,42%
14	Bestandteile – Beton	1,9365	2884,00	362,50	0,0185	60,00%	93,33%
15	Korrosion – pH – Stahl	1,3847	2342,67	298,33	0,0198	75,00%	100,00%
16	Beton – Frost	2,0330	2681,50	338,00	0,0224	80,00%	80,00%
17	Beton – Stahl	7,0130	4580,50	673,50	0,0227	50,00%	59,09%
18	Kunststoff – Polymerisation	0,7270	1494,00	198,50	0,0245	71,43%	100,00%
19	Legierung – Stahl	3,3780	2937,50	442,00	0,0260	60,00%	85,00%
20	Herstellung – Hochofen – Roheisen – Stahl	1,2848	1896,00	259,00	0,0262	80,00%	80,00%
21	Kunststoff – Herstellung	2,7010	2490,50	316,50	0,0343	66,67%	80,00%
22	Edelstahl – Stahlbau – Verwendung	0,2090	812,00	74,33	0,0346	50,00%	71,43%
23	Beton – Rohdichte – Zuschlag	2,2717	2224,67	273,67	0,0373	78,95%	84,21%
24	Beton – Bewehrung – Korrosion – Stahl	4,7685	3197,50	391,25	0,0381	56,52%	56,52%
25	Beton – Kriechen – Schwinden	2,0592	2014,00	252,00	0,0406	57,14%	64,29%
26	Aluminium – Verwendung	1,0530	1606,50	155,00	0,0423	81,82%	90,91%
27	Gleichgewichtsfeuchte – Holz – Quellen	1,3103	1737,00	167,00	0,0452	90,00%	100,00%
28	Beton – Verwendung – Zusatzmittel	3,0018	2334,33	278,33	0,0462	80,00%	80,00%
29	Abbinden – Zement	0,9040	1252,00	144,00	0,0501	50,00%	58,33%
30	Beton – Druckfestigkeit – Prüfung	3,1107	2174,00	264,00	0,0542	90,00%	90,00%
31	Beton – Druckfestigkeit	6,3550	2821,50	380,50	0,0592	72,22%	88,89%
32	Beton – Bewehrung – Korrosion	5,0013	2822,00	290,67	0,0610	66,67%	66,67%
33	Dauer – Erhärten – Frischbeton	0,3817	1040,67	59,33	0,0618	44,44%	100,00%

Nr.	Suchbegriffe	S	K	h	VS	P	P'
34	Beton – Druckfestigkeit – Zugfestigkeit	4,7237	2218,67	288,67	0,0738	68,18%	77,27%
35	Beton – Zusatzmittel	6,7740	2664,00	338,50	0,0751	78,57%	85,71%
36	Schweißen – Verfahren	3,6120	2195,50	193,00	0,0852	73,68%	84,21%
37	Feuchte – Holz – Quellen	4,0873	2239,67	185,67	0,0983	70,59%	82,35%
38	Herstellung – Hochofen – Roheisen	1,3453	1086,67	114,33	0,1083	84,62%	84,62%
39	Abbindevorgang – Estrich – Klebstoff	0,4250	664,67	58,00	0,1102	85,71%	100,00%
40	Feuchte – Holz	8,5080	2945,00	261,00	0,1107	82,35%	88,24%
41	Korrosion – Metall	6,2600	2452,50	216,00	0,1182	82,35%	82,35%
42	Erhärten – Frischbeton	0,6230	681,50	75,50	0,1211	58,33%	100,00%
43	Bewehrung – Korrosion – pH	2,3363	1358,33	97,33	0,1767	78,95%	94,74%
44	Bestandteil – Zement	4,7520	1501,50	164,50	0,1924	78,57%	100,00%
45	Elastisch – Plastisch – Verhalten	5,7783	1424,00	144,00	0,2818	90,00%	95,00%
46	Hydratation – Zement	6,6310	1334,50	154,50	0,3216	90,91%	100,00%
47	Anwendung – Laubholz – Nadelholz	1,9763	890,00	64,67	0,3434	77,78%	100,00%
48	Druckfestigkeit – Zugfestigkeit	4,5360	909,50	106,00	0,4705	87,50%	100,00%
49	Hochofen – Stahlerzeugung	0,3610	282,50	19,50	0,6553	77,78%	88,89%
50	Legierungselement – Mangan	1,5410	501,00	44,50	0,6912	78,57%	92,86%
51	Asphalt – Bitumen	1,8630	638,50	40,50	0,7204	72,73%	90,91%
52	Frost – Luftporen	0,7350	360,00	17,50	1,1667	71,43%	85,71%
53	Duroplast – Thermoplast	3,6150	609,50	46,00	1,2894	78,57%	100,00%
54	Elastomer – Thermoplast	4,3630	718,00	46,00	1,3210	76,47%	94,12%
55	Duroplast – Elastomer – Thermoplast	4,1923	651,67	44,67	1,4402	80,00%	100,00%
56	Löten – Flussmittel	3,0150	438,00	46,00	1,4964	92,86%	92,86%
57	duktil – Versagen	3,4590	436,50	49,50	1,6009	84,62%	84,62%
58	Polykondensation – Polymerisation	1,9700	268,00	23,50	3,1280	90,91%	90,91%
59	Polyaddition – Polymerisation	2,7920	254,50	19,50	5,6259	91,67%	100,00%
60	Polyaddition – Polykondensation – Polymerisation	2,7343	232,33	20,33	5,7881	83,33%	91,67%
61	Polyaddition – Polykondensation	3,4410	174,50	18,00	10,9551	90,00%	90,00%
62	flächenzentriert – kubisch	4,8180	114,50	20,00	21,0393	85,71%	85,71%
S: Signifikanzparameter der Suchbegriffe in der Suchbegriffskombination K: mittlere Kookkurrentenanzahl der Suchbegriffe in der Suchbegriffskombination h: mittlere Häufigkeit der Suchbegriffe in der Suchbegriffskombination VS: Verknüpfungsstärke der Suchbegriffe, $VS = (S \cdot 10.000) / (K \cdot h)$ P: prozentualer Anteil relevanter Wissenseinheiten im Lehrpfad P': prozentualer Anteil relevanter und inhaltlich ähnlicher Wissenseinheiten im Lehrpfad							

Das Streudiagramm der Werte VS und P ist in Abbildung 7.3 dargestellt. Jede untersuchte Suchbegriffskombination ist hier mit ihrem VS-Wert auf der logarithmisch skalierten Abszisse und ihrem P-Wert auf der Ordinate verzeichnet. Die ansteigende Tendenz der Precision für eine größer werdende Verknüpfungsstärke VS ist deutlich zu erkennen.

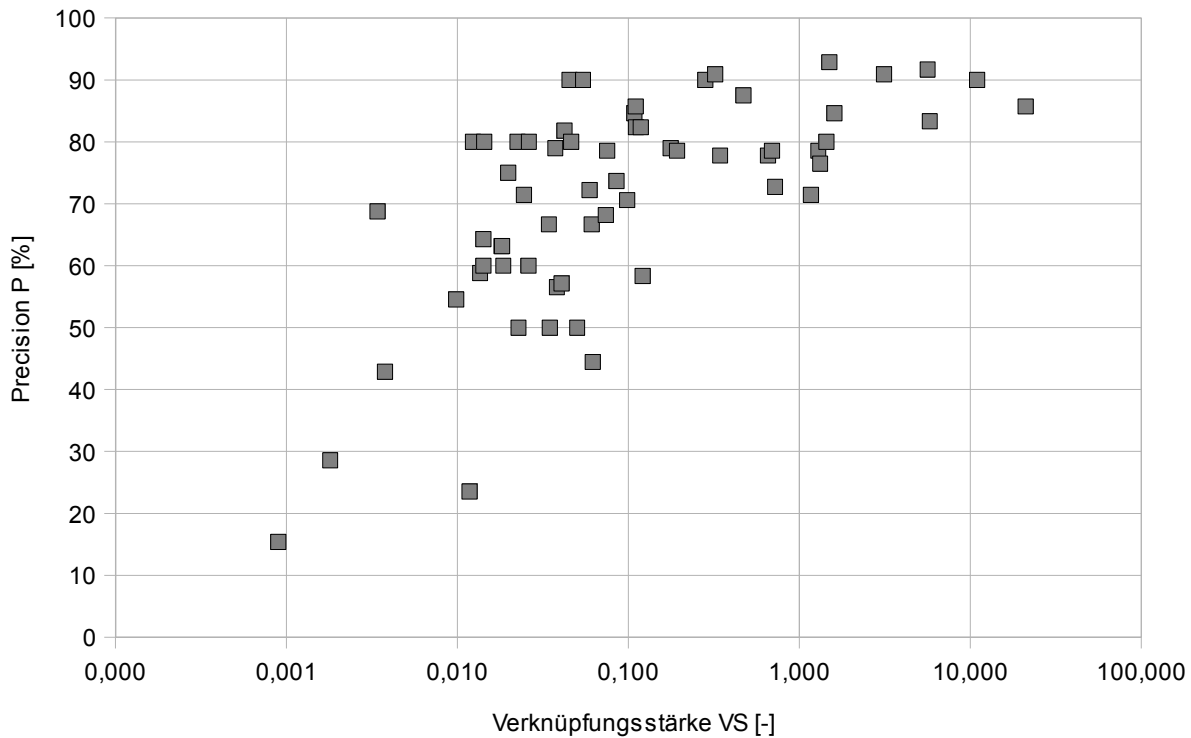


Abbildung 7.3: Verknüpfungsstärke der Suchbegriffe und Precision der daraus erstellten Lehrpfade für die untersuchten Suchwortkombinationen

Für die Bestimmung des Korrelationskoeffizienten bietet sich aufgrund der Verteilung der Werte eine logarithmische Regression in der Form $y = a + b \cdot \ln(x)$ an. Zunächst wird diese Formel linearisiert, indem für $\ln(x)$ w substituiert wird: $y = a + b \cdot w$. Für den empirischen Korrelationskoeffizienten r gilt dann:

$$r = \frac{n \cdot \sum w_i y_i - \sum w_i \cdot \sum y_i}{\sqrt{\left[n \cdot \sum w_i^2 - (\sum w_i)^2 \right] \cdot \left[n \cdot \sum y_i^2 - (\sum y_i)^2 \right]}} \quad (7.2)$$

Die entsprechenden Summen finden sich tabellarisch ermittelt in Anhang K. Es ergibt sich damit für $n = 62$ als Stichprobenumfang der Korrelationskoeffizient $r = 0,655$. Für die Parameter b („empirischer Regressionskoeffizient“) und a gilt entsprechend:

$$b = \frac{n \cdot \sum w_i y_i - \sum w_i \cdot \sum y_i}{n \cdot \sum w_i^2 - (\sum w_i)^2} = 5,20 \quad (7.3)$$

$$a = \frac{\sum y_i \cdot \sum w_i^2 - \sum w_i \cdot \sum w_i y_i}{n \cdot \sum w_i^2 - (\sum w_i)^2} = 83,49 \quad (7.4)$$

Aus dem für die Stichprobe ermittelten Korrelationskoeffizienten r kann der Korrelationskoeffizient ρ für die Grundgesamtheit abgeschätzt werden. Der Korrelationskoeffizient r wird als Realisation der Zufallsgröße R betrachtet [WEBER 1992]. Die Zufallsgröße R ist nicht normalverteilt wenn $\rho \neq 0$ gilt, was hier voraus zu setzen ist. Deshalb muss zunächst aus R mit der Fischerschen z -Transformation eine approximativ normalverteilte Zufallsgröße Z erzeugt werden:

$$Z = \operatorname{arctanh} R = \frac{1}{2} \cdot \ln \frac{1+R}{1-R} \quad (7.5)$$

Für deren Mittelwert μ und Varianz σ^2 gilt:

$$\mu = \frac{1}{2} \cdot \ln \frac{1+\rho}{1-\rho} + \frac{\rho}{2 \cdot (n-1)} \quad (7.6)$$

$$\sigma^2 = \frac{1}{n-3} \quad (7.7)$$

Da die Parameter μ und σ der Zufallsgröße Z bekannt sind, kann man Z in die standardnormalverteilte Zufallsgröße U transformieren, für die gilt:

$$U = \frac{Z - \mu}{\sigma} = (Z - \mu) \cdot \sqrt{n-3} \quad (7.8)$$

Mit dieser Größe kann jetzt gezeigt werden, dass ρ auf dem Signifikanzniveau $\alpha = 0,05$ größer als 0,5 ist. Die Nullhypothese H_0 lautet also $\rho \leq 0,5$. Die entsprechende Alternativ-Hypothese ist $\rho > 0,5$. Für den kritischen Bereich gilt, dass u als Realisation der Zufallsgröße U betragsmäßig größer sein muss als $u_{0,95} = 1,645$. Für u gilt:

$$u = (z - \mu_z) \cdot \sqrt{n-3} = \left(\frac{1}{2} \cdot \ln \frac{1+0,656}{1-0,656} - \frac{1}{2} \cdot \ln \frac{1+0,5}{1-0,5} + \frac{0,5}{2 \cdot (62-1)} \right) \cdot \sqrt{62-3} = 1,783 \quad (7.9)$$

Der Wert u liegt also im kritischen Bereich, deshalb kann die Nullhypothese verworfen werden. Der Korrelationskoeffizient ρ ist damit signifikant größer als 0,5. Es kann folglich von einer stochastischen Abhängigkeit der Precision von der Verknüpfungsstärke VS ausgegangen werden.

Im Folgenden soll ein Prognoseintervall bestimmt werden, mit dem mit einer bestimmten Wahrscheinlichkeit – im vorliegenden Fall $1 - \alpha = 0,90$ – die zu erwartende Precision eines Lehrpfads für den errechneten Wert VS_0 der entsprechenden Suchanfrage berechnet werden kann. Da die ursprüngliche Gleichung linearisiert wurde, muss statt mit VS_0 mit $\ln(VS_0)$ oder mit $w_0 = \ln(VS_0)$ gerechnet werden. In Anlehnung an HARTUNG ET AL. 1995 ist dieses Intervall

$$[a + b \cdot \ln(VS_0) - D; a + b \cdot \ln(VS_0) + D] \quad (7.10)$$

mit

$$D = s \cdot t_{n-2; 1-\frac{\alpha}{2}} \cdot \sqrt{1 + \frac{1}{n} + \frac{(\ln(VS_0) - \bar{w})^2}{\sum_{i=1}^n (w_i - \bar{w})^2}} \quad (7.11)$$

Dabei ist $t_{n-2; 1-\alpha/2} = 1,671$ das Quantil der t-Verteilung mit $n - 2 = 60$ Freiheitsgraden für das zweiseitige Signifikanzniveau $\alpha = 0,10$. Der Wert $s = 12,81$ ist ein Schätzwert für die Standardabweichung der Precision der Grundgesamtheit. Für die Berechnung von s gilt:

$$s = \sqrt{\frac{1}{n-2} \cdot \sum_{i=1}^n (P_i - a - b \cdot w_i)^2} = \sqrt{\frac{1}{n-2} \cdot \sum_{i=1}^n (P_i - P_{i0})^2} \quad (7.12)$$

P_{i0} stellt den Schätzwert für P an der Stelle VS_0 dar, es gilt:

$$P_{i0} = \bar{P} + b \cdot (w_0 - \bar{w}) \quad (7.13)$$

Der Bereich zwischen den Kurven durch die berechneten oberen und unteren Grenzen der Prognoseintervalle wird als Prognosestreifen bezeichnet. Er ist in Abbildung 7.4 dargestellt. Die berechneten Werte G_{oben} für die obere und G_{unten} für die untere Grenze des Intervalls finden sich in Anhang K.

Wesentlich schmaler als der Prognosestreifen ist der ebenfalls in Abbildung 7.4 dargestellte Konfidenzstreifen. Dieser gibt den wahrscheinlichen Bereich für die Größe des Erwartungswerts von P an einer Stelle VS_0 an. Der Konfidenzstreifen beruht auf den Konfidenzintervallen, die nach HARTUNG ET AL. 1995 analog zu den Prognoseintervallen bestimmt werden. Ein Konfidenzintervall an der Stelle VS_0 ist:

$$[a+b \cdot \ln(VS_0) - C; a+b \cdot \ln(VS_0) + C] \quad (7.14)$$

mit

$$C = s \cdot t_{n-2; 1-\frac{\alpha}{2}} \cdot \sqrt{\frac{1}{n} + \frac{[\ln(VS_0) - \bar{w}]^2}{\sum_{i=1}^n (w_i - \bar{w})^2}} \quad (7.15)$$

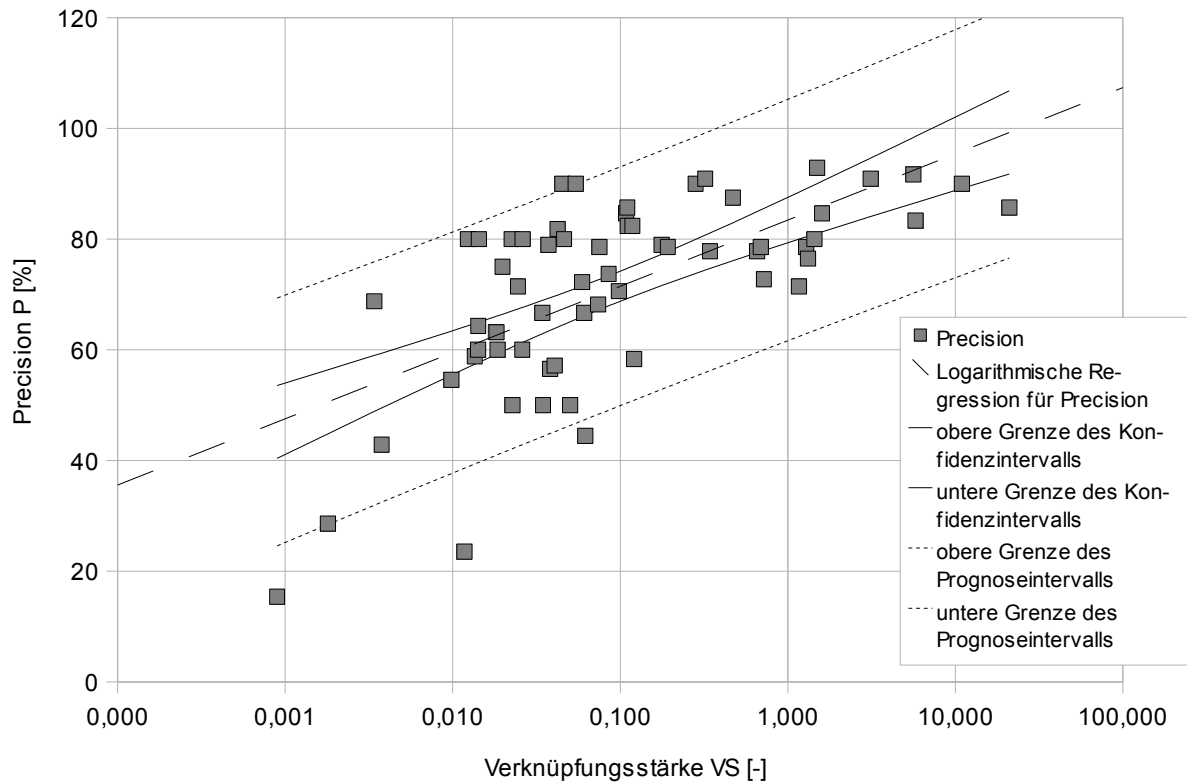


Abbildung 7.4: Zusammenhang zwischen Verknüpfungsstärke und Precision, dargestellt anhand des Konfidenz- und des Prognosebereichs

Für den unteren Grenzwert $P_{0, \text{unten}}$ des Prognosestreifens an einer Stelle VS_0 gilt:

$$P_{0, \text{unten}} = a + b \cdot \ln(VS_0) - s \cdot t_{n-2; 1-\frac{\alpha}{2}} \cdot \sqrt{1 + \frac{1}{n} + \frac{[\ln(VS_0) - \bar{w}]^2}{\sum_{i=1}^n (w_i - \bar{w})^2}} \quad (7.16)$$

die Formel vereinfacht sich zu

$$P_{0,unten} = \bar{P} + b \cdot [\ln(VS_0) - \bar{w}] - s \cdot t_{n-2; 1-\frac{\alpha}{2}} \cdot \sqrt{1 + \frac{1}{n} + \frac{[\ln(VS_0) - \bar{w}]^2}{\sum_{i=1}^n (w_i - \bar{w})^2}} \quad (7.17)$$

Anhand der beiden Intervalle kann die voraussichtliche Qualität eines Lehrpfades, ausgedrückt durch seine Precision, aus der Suchanfrage berechnet werden. Der jeweilige Wert $P_{0,unten}$ wird dem Lerner dann mitgeteilt. Für eine Verknüpfungsstärke von $VS_0 = 1,0$ ergibt sich somit ein unterer Grenzwert für die Precision des Lehrpfads von 61,64%. Dies ist scheinbar ein niedriger Wert, jedoch muss beachtet werden, dass der Mittelwert aller Lehrpfade, die denselben Wert für VS_0 aufweisen, schon zwischen 79,33% und 87,53% liegt. Ferner ist an die Lehrpfade ein strenges Qualitätskriterium angelegt worden, indem dem Lernziel nicht direkt zuzuordnende, sondern lediglich verwandte Seiten als irrelevant ausgeschlossen wurden. Allerdings wurde das eventuelle Fehlen von Inhalten nicht berücksichtigt, die tatsächliche Qualität des Lehrpfads kann also unter der dargestellten liegen.

Die zusätzliche Berücksichtigung zum Lernziel ähnlicher Wissenseinheiten wird im nächsten Abschnitt untersucht.

7.3 Relevanz der Lehrpfade unter Berücksichtigung lernzielähnlicher Seiten

Zum Vergleich mit den im vorangegangenen Abschnitt betrachteten Relevanzwerten, die nur die für das Lernziel des Lehrpfads absolut notwendigen Seiten berücksichtigen, soll an dieser Stelle untersucht werden, wie bei einem „weicheren“ Relevanzkriterium die Qualität der Lehrpfade zu beurteilen ist.

Es wurden deshalb alle Lehrpfade ein zweites Mal dahingehend beurteilt, ob diejenigen der enthaltenen Wissenseinheiten, die zuvor als irrelevant eingestuft wurden, ähnliche Inhalte zum Thema haben wie das Lernziel des Lehrpfads. Hierbei wurden wie gehabt nur die Haupt- und Nebenseiten betrachtet, weiterführende Seiten wurden nicht berücksichtigt, da sie unabhängig von der Zusammenstellung des Lehrpfads sind. Die Precision P' der auf diese Weise beurteilten Lehrpfade ist also immer größer oder gleich der ausschließlich auf das Lernziel bezogenen Precision P . Die Bestimmung der P' -Werte kann Anhang J entnommen werden, die ermittelten Wertepaare sind in dem Streudiagramm in Abbildung 7.5 dargestellt. Es ergibt sich ein etwas schlechterer Korrelationskoeffizient $r = 0,605$. Für diesen kann auf dem Signifikanzniveau $\alpha = 0,05$ ein Korrelationskoeffizient von $p \geq 0,40$ für die Grundgesamtheit nachgewiesen werden. Eine stochastische Abhängigkeit ist also auch hier eindeutig vorhanden.

Der Konfidenz- und der Prognosestreifen werden analog zum Vorgehen in Abschnitt 7.2 berechnet. Man sieht in Abbildung 7.5 deutlich die auf höherem Niveau gelegenen Intervallgrenzen. Hier wird die Untergrenze für eine Precision $P' = 60\%$ schon bei $VS_0 = 0,0454$ erreicht. Ab ungefähr demselben Wert können – an der Obergrenze des Prognoseintervalls – auch P' -Werte von 100% erreicht werden. Für $VS_0 = 1,0$ wird eine Mindest-Precision von 72,85% erreicht. Nach analoger Rechnung ergibt sich die untere Grenze des Konfidenzintervalls, also des Bereichs, in dem der Mittelwert für mehrere Suchanfragen mit demselben VS-Wert erwartet werden kann, für $VS_0 = 0,1$ zu $P_0' = 80,85\%$.

Der jeweils berechnete „weiche“ Relevanzwert kann dem Lerner ergänzend angegeben werden, wenn er dies wünscht.

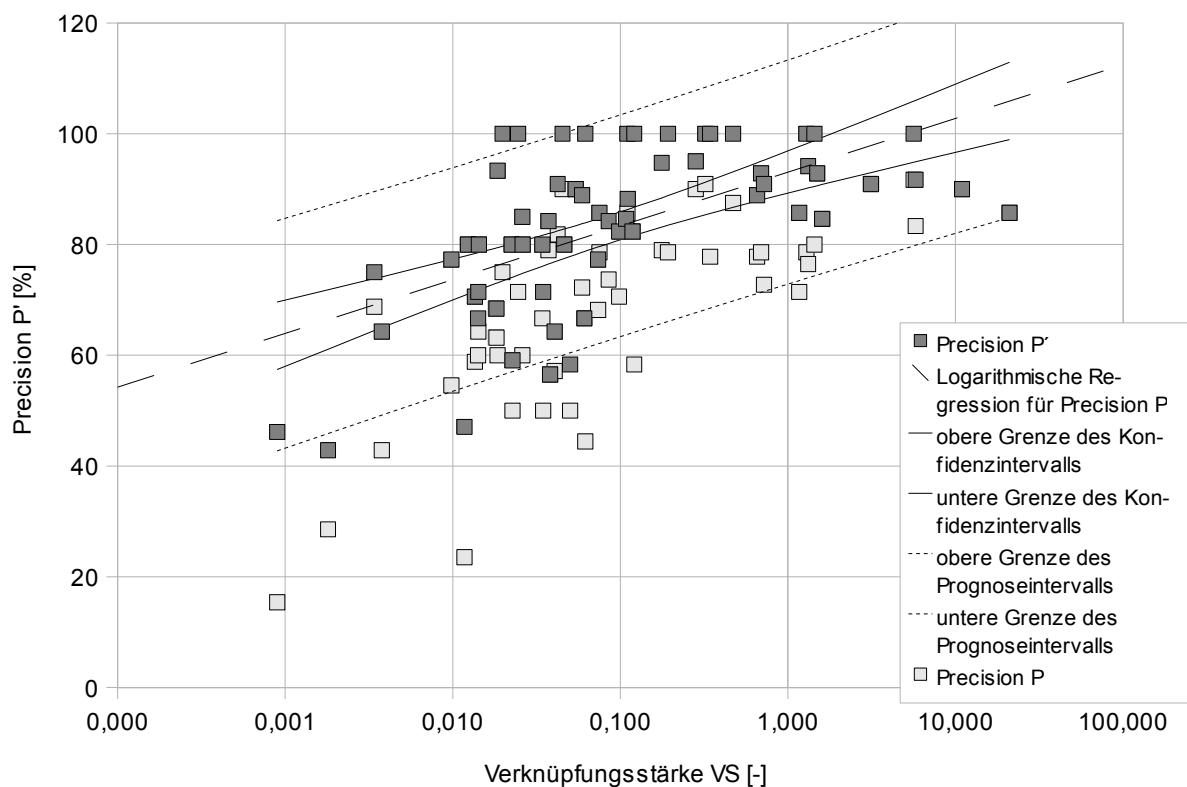


Abbildung 7.5: Verknüpfungsstärke und Precision P' unter Berücksichtigung lernzielverwandter Wissensseinheiten
(ursprüngliche Precision-Werte P sind zum Vergleich eingezeichnet)

7.4 Schlussfolgerungen

Im Folgenden soll diskutiert werden, wie sich die Auswertung der Lehrpfade hinsichtlich inhaltlicher Relevanz, didaktischer Struktur und Erweiterbarkeit des Systems interpretieren lässt.

Die Untersuchungen an 62 Lehrpfaden zeigen, dass allgemein von einer guten inhaltlichen Qualität der Lehrpfade für den Einsatz in der Praxis ausgegangen werden kann. Diese Qualität ist für spezielle Themen, deren Suchbegriffe semantisch verwandt sind, höher als für weit gefasste Themengebiete. Das System zeigt dem Lerner auf einem Signifikanzniveau $\alpha = 0,05$ an, wie hoch die Precision eines von ihm erzeugten Lehrpfads mindestens ist.

Für Lernende im Grundstudium empfehlen sich rudimentäre Vorkenntnisse der Thematik, damit Suchbegriffe richtig angewendet und die in ersten selbst generierten Lehrpfaden enthaltenen Inhalte interpretiert werden können. Der Beginn der Nutzung des vorgeschlagenen Systems ist deswegen erst nach einigen einführenden Vorlesungen zu empfehlen.

Für Studierende im Hauptstudium, die Zusammenhänge wiederholen oder vertiefen wollen, ist das System ohne Einschränkungen nutzbar. Diese Lerner verfügen nach bestandem Vordiplom prinzipiell über ein höheres Vorwissen. Jedoch besteht nach STEINBORN 2007 bei Studierenden die Tendenz, weniger Wert auf Hintergrundwissen als auf eine bestandene Prüfung zu legen. Eventuelle Lücken im Fachwissen könnten also mittels individuell erstellter Lehrpfade einfach geschlossen werden.

Lehrende können die Qualität der erzeugten Lehrpfade selbst beurteilen. Für sie ist ein Hinweis auf die wahrscheinliche Precision eines für die Studierenden erzeugten Lehrpfads folglich entbehrlich. Sie können jedoch durch eine Auswahl von solchen Suchbegriffen, die semantisch stark miteinander verknüpft sind, von vornherein Lehrpfade mit höherer Relevanz erstellen und damit Aufwand für eine redaktionelle Überarbeitung einsparen.

Neben den Inhalten der Lehrpfade ist auch deren Aufbau wichtig. Da eine entsprechende objektive Bewertung schwierig ist, wurde auf eine diesbezügliche Auswertung der 62 untersuchten Lehrpfade verzichtet. Es können aber die in Abschnitt 7.2 exemplarisch betrachteten Lehrpfade diskutiert werden.

Allgemein sieht man, dass die Lehrpfade gut nach ähnlichen Inhalten gruppiert werden. Dies fällt vor allem bei den nicht zum eigentlichen Thema passenden Seiten auf (z.B. Klebstoffe beim Abbinden von Zement und Stromquellen beim Quellen von Holz). Das Clustering ist hinreichend stabil, wird also nicht bei jeder Neukombination von Seiten bei einer Änderung der Suchanfrage anders durchgeführt. Dies wurde anhand von Lehrpfad 2 und seinen Alternativen gezeigt. Jedoch können zusätzliche Seiten auch die Bildung anders zusammengestellter Cluster verursachen. In der Regel enthalten Lehrpfade, die einen Anteil derselben Seiten enthalten, auch ähnliche aus diesen gebildete Cluster. Durch die Anhebung des Grenzwertes G_C für die erste einem Clusterkern zuzufügende Wissensseinheit und dessen spätere Absenkung kann eine ausgewogene Verteilung der Clusterrumfänge erreicht werden.

Das System leistet die Bestimmung einer Reihenfolge für die Cluster nach implizierten didaktischen Eigenschaften der enthaltenen Wissenseinheiten. Leider können sich diese Eigenschaften mehrerer Seiten in einem Cluster gegenseitig beeinflussen, da die Cluster nur nach Ähnlichkeitskriterien gebildet werden. Dem Lerner wird deshalb zur Orientierung das Ordnungsprinzip der ausgewählten Reihenfolge angezeigt. Hauptseiten und Nebenseiten werden nach dem Grundprinzip, die allgemeinste Seite als Hauptseite zu verwenden, ausgewählt. Hier ist zu beachten, dass die Auswahl einer von mehreren hierarchisch gleichberechtigten Seiten als übergeordnete Seite zu willkürlichen Ergebnissen führen kann.

Im Hinblick auf die statistische Auswertung und die Prognose der Lehrpfadqualität stellt sich die Frage, wie das System sich bei einer Erweiterung verhält. Der gezeigte Zusammenhang zwischen Verknüpfungsstärke VS und Precision P ist mit den angegebenen Zahlenwerten nur für die Inhaltsbasis des WiBA-Net im aktuellen Zustand gültig. Der Wert VS lässt sich zwar auch für Suchanfragen auf anderen Inhaltsbasen berechnen. Wegen der nicht linear zusammenhängenden Größen S, K und h werden sich jedoch für dieselbe Suchanfrage bei verschiedenen Inhaltsbasen wegen des unterschiedlichen Umfangs der Textkorpora und der unterschiedlichen Worthäufigkeiten auch unterschiedliche Werte für VS ergeben. Folglich ist beim Einsatz einer neuen Inhaltsbasis mit dem vorgeschlagenen System eine erneute Regressionsrechnung nach dem in Abschnitt 7.3 angegebenen Muster durchzuführen.

Aus denselben Gründen ist der anhand der Zahlenwerte aus Anhang K für die Inhalte des WiBA-Net ermittelte Zusammenhang auch nur gültig, bis eine bestimmte Anzahl neuer Wissenseinheiten in das System aufgenommen wurde. Es sind also Neukalibrierungen nach bestimmten Erweiterungsumfängen notwendig, wenn die Lehrpfadqualität auch bei einer vergrößerten Inhaltsbasis korrekt prognostiziert werden soll. Diese Notwendigkeit bezieht sich wohlgerneht nur auf die Prognose der Qualität, nicht auf die Qualität selber. Diese kann sich durch eine Vergrößerung der Inhaltsbasis und die damit einhergehende stärkere Vernetzung der einzelnen Begriffe sogar verbessern. Ein Kriterium zur Bestimmung des Zeitpunktes der Neukalibrierung könnte über eine Speicherung berechneter VS-Werte definiert werden. Diese würden bei inhaltlichen Erweiterungen des Systems neu berechnet und mit den alten Werten verglichen werden. Bei einer festzulegenden durchschnittlichen Abweichung der alten von den neuen Werten müsste dann eine Neukalibrierung durchgeführt werden.

8 Zusammenfassung und Ausblick

8.1 Zusammenfassung der Ergebnisse

Im Rahmen der vorliegenden Arbeit wurden die theoretischen Grundlagen für ein System zur automatischen Erstellung von Lehrpfaden entwickelt und ein Prototyp dieses Systems implementiert. Inhaltlich fußt das System auf dem Lernnetzwerk „Werkstoffe im Bauwesen“ (WiBA-Net), in dessen didaktisches Konzept es auch integriert wurde.

Es wurde bewusst auf die Nutzung von Metadaten, manuellen Verknüpfungen zwischen Begriffen oder Kategorisierungen von Wissensseinheiten oder Begriffen verzichtet. Auf diese Weise sollte die Leistungsfähigkeit eines E-Learning-Systems überprüft werden, das ausschließlich aus dem Text extrahierbare Informationen zur Zusammenstellung von Lerneinheiten nutzt.

Die Inhalte des WiBA-Net wurden für das System aufbereitet, indem linguistische und textstatistische Informationen aus dem Textkorpus gewonnen wurden. Diese werden zur Suche, inhaltlichen Auswahl und didaktisch sinnvollen Zusammenstellung von Wissensseinheiten auf der Basis einer Suchanfrage genutzt.

Es hat sich gezeigt, dass die in den erstellten Lehrpfaden enthaltenen Wissensseinheiten von unterschiedlicher Relevanz für das durch die Suchanfrage implizierte Lern- oder Lehrziel des Nutzers sind. Es konnte gezeigt werden, dass diese unterschiedliche Relevanz zu einer neu definierten Kenngröße des Textkorpus, der Verknüpfungsstärke, korreliert ist. Aus dieser Korrelation konnte eine Möglichkeit der Prognose der Relevanz der Lehrpfadinhalte abgeleitet werden.

Diese Prognose wird den Lernern mitgeteilt, damit diese ihr Lernverhalten auf die zu erwartende Relevanz des für sie erstellten Lehrpfads abstimmen können.

8.2 Didaktischer Mehrwert des Systems

Der Vorteil des entwickelten Systems gegenüber anderen E-Learning-Systemen ist die Zusammenstellung von Lerneinheiten in didaktisch strukturierter Form speziell auf eine Nutzeranfrage hin. Das in das didaktische Konzept des WiBA-Net eingepasste Einsatzkonzept des Systems sieht vor, kein völlig eigenständiges Lernen mit dem System durchzuführen, sondern es im Rahmen einer universitären Lehrveranstaltung parallel zu Präsenzveranstaltungen zu nutzen.

Durch diese in das Konzept des Blended Learning einzuordnende Vorgehensweise muss sich der Lerner – insbesondere, solange er noch keinerlei Fachkenntnisse hat – nicht alleine in das Thema „Werkstoffe im Bauwesen“ einarbeiten. Er erhält stattdessen zunächst grundlegende, später hauptsächlich vertiefende Informationen in der Vorlesung und kann das System angeleitet zur Vor- und

Nachbereitung nutzen. Wenn die Kenntnisse eines Lerners fortgeschritten sind, kann er es auch nutzen, um nach eigenen Interessen sein Wissen zu vertiefen oder dieses Wissen im Rahmen einer Aufgabenstellung eigenständig zu erarbeiten. Da das System auf individuelle Anfragen Lerneinheiten erzeugt, können auch z.B. Projektaufgaben sinnvoll unterstützt werden. Die benötigten Informationen sind zentral vorhanden und können durch einen fortschreitenden Kenntnisstand eigenständig erschlossen werden.

Gegenüber einem reinen Nachschlagewerk hat das System den Vorteil, dass die durch die Suchanfrage implizierten Inhalte immer in ihrem Kontext angezeigt werden. Sind Inhalte für eine Suchanfrage relevant, so werden sie auch in den entsprechenden Lehrpfad einbezogen. Sie können also in verschiedenen Kontexten auftreten. Der Lerner erhält so verschiedene Blickwinkel auf das Gelernte. Dadurch implizierte Zusammenhänge können vom Lerner selbst konstruiert werden, das „Verständnis des Ganzen“ wird gefördert. Durch die Einbindung von Inhalten in ihren Kontext werden für den Lerner außerdem auch Themen erschlossen, die ihm noch nicht bekannt sind und nach denen er folglich auch nicht suchen würde.

Das System ist insofern adaptiv, als es sich an die individuellen, aktuellen Lernziele der Studierenden anpasst und keine feststehenden Lerneinheiten angeboten werden. Als Aspekt der Diversität der Lernenden können deren unterschiedliche Interessen oder auch Wissenslücken berücksichtigt werden, auch wenn ihr konkreter Wissensstand dem System nicht bekannt ist. Auf diese Weise kann der Lerner dem System seinen Informationswunsch vorgeben, anstatt vom System auf der Basis seines (vermuteten) Kenntnisstandes eine passende Lerneinheit präsentiert zu bekommen. Diese Vorgehensweise kann nach Bedarf mit den feststehenden Lehrpfaden des WiBA-Net kombiniert werden.

Durch die Einbindung des Systems in die Vorlesung und die individuelle Inhaltserstellung auch für Hochschullehrer werden auch diese didaktisch weitgehend unterstützt, indem es ihnen ermöglicht wird, die von ihnen gelehrteten Schwerpunktsthemen in selbst erstellten E-Learning-Einheiten den Studierenden zugänglich zu machen. Durch die Auslagerung der Vorlesungsinhalte in die durch E-Learning unterstützten Vor- und Nachbereitungsphasen entstehen außerdem Freiheitsgrade in der Vorlesung, die nach eigenem Ermessen des Hochschullehrers genutzt werden können.

8.3 Ausblick

Die Ergebnisse der vorliegenden Arbeit sowie die erkannten Schwachpunkte der automatischen Lehrpfadgenerierung geben Anlass, einen Ausblick auf Entwicklungsmöglichkeiten des Systems zu geben.

Ein in dieser Arbeit nicht berücksichtigter Aspekt ist die Anwendung einer Ontologie bzw. Taxonomie, um die Beziehungen zwischen Wissensseinheiten explizit speichern und nutzen zu können. Dies umfasst im weiteren Sinn auch die Nutzung manuell erstellter oder zumindest kontrollierter Wörterbücher und Thesauren. Die Nutzung verifizierter Beziehungen zwischen Begriffen erleichtert die Auswahl von Wissensseinheiten, da auf diese Weise nicht nur der Kontext der enthaltenden Begriffe eine Aussage über die Ähnlichkeit zu anderen Wissensseinheiten zulässt. Auf diese Weise ließe sich z.B. die in Abschnitt 5.3.5 beschriebene fälschliche Einbindung von Wissensseinheiten, die Kohyponyme von Suchbegriffen thematisieren, vermeiden. Auch Haupt- und Nebenseiten könnten mit Informationen über die hierarchische Anordnung der entsprechenden Wissensseinheiten einfacher festgelegt werden. Eine solche Ontologie muss zunächst in geeigneter Form erstellt und dann bei einer inhaltlichen Erweiterung des Systems ergänzt werden. Unterstützende Verfahren, auch auf Basis der automatischen Sprachverarbeitung, sind aber vorhanden (vgl. z.B. [HEYER ET AL. 2006]). Ein weiterer Vorteil der Entwicklung einer solchen Ontologie ist der dann auch mögliche Einsatz mehrsprachiger Inhalte, wenn (Fach-)Terme in fremden Sprachen einander zugeordnet werden.

Auch die Metadaten der WiBA-Net-Inhalte wie Stichworte und Beschreibungen von Inhalten wurden nicht genutzt. Diese könnten bei der Booleschen Suche nach Wissensseinheiten sinnvoll eingesetzt werden.

Die Benutzerschnittstelle ermöglicht die Eingabe von Suchbegriffen, deren Kontext das System interpretiert. Eine natürlichsprachliche Schnittstelle, die das genaue Ziel der Fragestellung erkennt und auch Verknüpfungen zwischen den Suchtermen berücksichtigt, würde das System insbesondere für Personen ohne Vorkenntnisse leichter bedienbar machen. Diese Entwicklung hat auch Bezüge zum o.a. Einsatz einer Ontologie.

Ein weiterer Gesichtspunkt ist die Weiterentwicklung der didaktischen Aspekte. So lässt sich ein System wie das vorgestellte auch im Rahmen kooperativen Lernens, z.B. bei Projektarbeiten, einsetzen. Lerner könnten so benötigtes Wissen, auch unter Zuhilfenahme des Systems, selbst konstruieren, mit dem Autorenwerkzeug formalisieren und anderen Lernern zugänglich machen. Auch ein Feedback für den Lerner würde das System didaktisch sinnvoll ergänzen. Dies könnte z.B. realisiert werden, indem den einzelnen Wissensseinheiten Testfragen zugeordnet werden, die dann im Anschluss an einen generierten Lehrpfad dem Lerner ebenfalls angezeigt werden.

9 Literaturverzeichnis

- Abicht, Lothar; Dubiel, Gerald: "E-Learning in der beruflichen Weiterbildung", in: Lernen und Weiterbildung als permanente Personalentwicklung - Band 1 (Hrsg: Peters, S.), Rainer Hampp-Verlag, München, 2003
- Ackermann, Martin: "Statistische Korpusanalyse zum Extrahieren von semantischen Wortrelationen", Dissertation, Universität Hildesheim, Institut für Mathematik und Angewandte Informatik, Hildesheim, 2000
- Advanced Distributed Learning Initiative: SCORM (2008), <http://www.adlnet.gov/scorm.aspx> (nachgewiesen am 02.06.2008)
- Apel, Hans Jürgen: Die Vorlesung - Einführung in eine akademische Lehrform, Böhlau, Köln, 1999
- Aroyo, Lora; Dicheva, Darina: "The New Challenges for E-Learning: The Educational Semantic Web", in: Educational Technology & Society, Band 7, 2004, S. 59-69
- Baeza-Yates, Ricardo; Ribeiro-Neto, Berthier: Modern Information Retrieval, ACM Press, New York, 1999
- Bayerischer Rundfunk: Telekolleg MultiMedial (2008), <http://www.br-online.de/wissen-bildung/telekolleg/> (nachgewiesen am 19.06.2008)
- Berlin-Brandenburgische Akademie der Wissenschaften: Digitales Wörterbuch der deutschen Sprache des 20. Jahrhunderts (2008), <http://www.dwds.de> (nachgewiesen am 29.05.2008)
- Berners-Lee, Tim; Hendler, James; Lassila, Ora: The Semantic Web (2001), <http://www.sciam.com/article.cfm?id=the-semantic-web> (nachgewiesen am 26.06.2008)
- Bloom, Benjamin S. (Hrsg.): Taxonomy of Educational Objectives. The classification of educational goals (Handbook I. Cognitive Domain), Longman, New York, 1956
- Borrmann, Jens; Schwarte, Joachim: "Einsatz von LORe (Learning Objects Repository) am Institut für Werkstoffe im Bauwesen der Universität Stuttgart", in: Neue Medien in der Aus- und Weiterbildung von Bauingenieuren und Architekten (Tagungsband) (Hrsg: Grübl, Peter; Klaus, Hans G.), Eigenverlag, Darmstadt, 2003
- Bremer, Claudia: "E-Learning-Strategien im Spannungsfeld von Hochschulentwicklung, Kompetenzansätzen und Anreizsystemen", in: E-Learning-Strategien und E-Learning-Kompetenzen an Hochschulen (Hrsg: Bremer, Claudia; Kohl, Kerstin E.), W. Bertelsmann Verlag, Bielefeld, 2004
- Brin, Sergey; Page, Lawrence: The Anatomy of a Large-Scale Hypertextual Web Search Engine (1998), <http://www-db.stanford.edu/~backrub/google.html> (nachgewiesen am 20.06.2008)

- Chang, Kenneth: NCSA Mosaic Exploits Internet (1993), <http://www.ncsa.uiuc.edu/News/Access/Archive/backissues/93.2/Mosaic.html> (nachgewiesen am 30.05.2006)
- Della Corte, Francesco: "E-Learning as educational tool in emergency and disaster medicine teaching", in: *Minerva Anesthesiologica*, Band 71, 2005, S. 181-195
- Dichanz, Horst: "Psychologie lehren - Psychologie lernen? Überlegungen zum E-Learning: Begriffe, Konzepte und Erfahrungen", in: *Psychologie Lehren und Lernen. Beiträge zur Hochschuldidaktik* (Hrsg: Steinebach, Christoph), Universitätsverlag Winter, Heidelberg, 2005
- Directorate-General for Education and Culture: ECTS Users' Guide - European Credit Transfer and Accumulation System and the Diploma Supplement (2005), http://www.hrk.de/de/download/dateien/ECTSUsersguide_en_Februar2005.pdf (nachgewiesen am 18.06.2008)
- Europäische Bildungsminister: Der Europäische Hochschulraum - Gemeinsame Erklärung der Europäischen Bildungsminister (1999), http://www.bmbf.de/pub/bologna_deu.pdf (nachgewiesen am 18.06.2008)
- Ferber, Reginald: *Information Retrieval - Suchmodelle und Data-Mining-Verfahren für Textsammlungen und das Web*, dpunkt, Heidelberg, 2003
- Fischer, Mario: *Website Boosting - Suchmaschinen Optimierung, Usability, Webseiten-Marketing*, mitp, Heidelberg, 2006
- Gage, Nathaniel L.; Berliner, David C.: *Pädagogische Psychologie*, Beltz, Weinheim, 1986
- Grübl, Peter; Reinhardt, Hans-Wolf; Setzer, Max Josef; Franke, Lutz; König, Gerd; Hillemeier, Bernd: *WiBA-Net - Werkstoffe im Bauwesen* (2004), <http://www.wiba-net.de> (nachgewiesen am 22.01.2006)
- Grübl, Peter; Schnittker, Nils; Schmidt, Bernd: "Gibt es den "elektronischen Nürnberger Trichter"? - Das Konzept des Blended Learning, dargestellt am Lernnetz WiBA-Net", in: *Digitaler Campus - Vom Medienprojekt zum nachhaltigen Medieneinsatz in der Hochschule* (Hrsg: Kerres, Michael; Voß, Britta), Waxmann, Münster, 2003
- Häfele, Hartmut; Maier-Häfele, Kornelia: *101 e-learning Seminarmethoden - Methoden und Strategien für die Online- und Blended Learning Seminarpraxis*, managerSeminare Verlags GmbH, Bonn, 2004
- Hagemann, Hans-Jürgen: "INGMEDIA: Lernsoftware für technische und physikalische Praktika in Ingenieurstudiengängen", in: *E-Learning - NMB-Projekte in den Ingenieurwissenschaften* (Hrsg: Pinkau, Stephan; Gerke, Thomas), Eigenverlag, Dessau, 2003
- Hartung, Joachim; Elpelt, Bärbel; Klösener, Karl-Heinz: *Statistik - Lehr- und Handbuch der angewandten Statistik*, R. Oldenbourg Verlag, München, 1995

- Hessischer Rundfunk: Funkkolleg (2008), <http://www.hr-online.de/website/radio/hr2/index.jsp?rubrik=2884> (nachgewiesen am 19.06.2008)
- Heyer, Gerhard; Läuter, Martin; Quasthoff, Uwe; Wittig, Thomas; Wolff, Christian: "Learning Relations using Collocations", in: Proceedings of the IJCAI Workshop on Ontology Learning, Seattle, 2001
- Heyer, Gerhard; Quasthoff, Uwe; Wittig, Thomas: Text Mining: Wissensrohstoff Text - Konzepte, Algorithmen, Ergebnisse, W3L-Verlag, Herdecke, 2006
- Jacobs, Ian: About W3C: History (2006), <http://www.w3.org/Consortium/history.html> (nachgewiesen am 29.05.2006)
- Karagiannis, Dimitris; Telesko, Rainer: Wissensmanagement - Konzepte der Künstlichen Intelligenz und des Softcomputing, Oldenbourg, München, 2001
- Kerres, Michael: "Didaktische Konzeption multimedialer und telemedialer Lernumgebungen", in: HMD - Praxis der Wirtschaftsinformatik, Band 205, 1999, S. 9-17
- Kerres, Michael: Multimediale und telemediale Lernumgebungen - Konzeption und Entwicklung, Oldenbourg, München, 2001
- Kerres, Michael; de Witt, Claudia; Stratmann, Jörg: "E-Learning. Didaktische Konzepte für erfolgreiches Lernen", in: Jahrbuch Personalentwicklung & Weiterbildung 2003 (Hrsg. von Schwuchow, K.; Guttmann, J.), Luchterhand, Unterschleißheim, 2002
- Kerres, Michael; Euler, Dieter; Seufert, Sabine; Hasanbegovic, Jasmina; Voss, Britta: "Lehrkompetenz für eLearning-Innovationen in der Hochschule - Ergebnisse einer explorativen Studie zu Maßnahmen der Entwicklung von eLehrkompetenz", in: Swiss Centre for Innovation in Learning - Arbeitsbericht 6, St. Gallen, 2005
- Kindt, Michael: Portal zur BMBF-Förderung Neue Medien in der Bildung - Kursbuch (2006), http://www.medien-bildung.net/produkte/produkte_pdfs.php/hochschule/produkte/0/0/0/0/0/ (nachgewiesen am 28.05.2006)
- Köhler, Stefan: "Ein ingenieurgerechtes Modell zur elektronischen Wissensvermittlung im Bauwesen", Dissertation, Technische Universität Darmstadt, Fachbereich Bauingenieurwesen und Geodäsie, Darmstadt, 2006
- König, Werner: dtv-Atlas Deutsche Sprache, Deutscher Taschenbuch Verlag, München, 2001
- Lee, Lillian: "Measures of distributional similarity", in: 37th Annual Meeting of the Association for Computational Linguistics (ACL), College Park, 1999
- Lewandowski, Dirk: Web Information Retrieval - Technologien zur Informationssuche im Internet, DGI-Schrift (Informationswissenschaft 7), Frankfurt am Main, 2005

- Meibauer, Jörg; Demske, Ulrike; Geilfuß-Wolfgang, Jochen; Pafel, Jürgen; Ramers, Karl-Heinz; Rothweiler, Monika; Steinbach, Markus: Einführung in die germanistische Linguistik, Verlag J.B. Metzler, Stuttgart, 2002
- Menzel, Karsten; Pottgiesser, Uta: "Projektbasierte, teamorientierte Lehr-Lern-Strategien zum e-Learning im Bauwesen", in: Neue Medien in der Aus- und Weiterbildung von Bauingenieuren und Architekten (Tagungsband) (Hrsg: Grübl, Peter; Klaus, Hans G.), Eigenverlag, Darmstadt, 2003
- Möbius, Bernd: "Sprachsynthesysteme", in: Computerlinguistik und Sprachtechnologie - Eine Einführung (Hrsg: Carstensen, K.-U.; Ebert, Ch.; Endriss, C.; Jekat, S.; Klabunde, R.; Langer, H.), Spektrum Akademischer Verlag, Heidelberg, 2001
- Montandon, Corinne: "Adoption von Standardisierung im E-Learning - Eine Umfrage bei E-Learning-Projekten an Hochschulen im deutschen Sprachraum", in: Institut für Wirtschaftsinformatik der Universität Bern - Arbeitsbericht 180, Bern, 2006
- Mühlhäuser, Max: "eLearning After Four Decades: What About Sustainability?", in: Proceedings of ED-MEDIA 2004 (Hrsg: Association for the Advancement of Computers in Education (AACE)), , Charlottesville, 2004
- Münz, Stefan: HTML-Zeichenreferenz (2007), <http://de.selfhtml.org/html/referenz/zeichen.htm> (nachgewiesen am 01.06.2008)
- Papula, Lothar: Mathematik für Ingenieure und Naturwissenschaftler (Band 3), Vieweg, Braunschweig, 1994
- Paraskakis, Iraklis: Ambient Learning: a new paradigm for e-learning (2005), <http://www.forma-tex.org/micte2005/229.pdf> (nachgewiesen am 31.05.2006)
- Pfäffli, Brigitta K.: Lehren an Hochschulen - Eine Hochschuldidaktik für den Aufbau von Wissen und Kompetenzen, Haupt, Bern, 2005
- Rath, Ralf; Baumer, Claudia: "Intelligente Wissensarchitekturen - Informationsräume durch semantische Wissensnetze erschließen", in: Worauf sollen wir bauen? Neue Medien in der Architekturausbildung, Wismar, 2003
- Reinhardt, Hans-Wolf (Hrsg.): "Hochschullehrermemorandum: Werkstoffe im Bauwesen - Universitäre Lehre und Forschung", in: Bauingenieur, Band 75 (11), 2000, S. 723-729
- Röhrig, Christof; Jochheim, Andreas: "The Virtual Lab for Teleoperated Control of Real Experiments", in: IEEE Conference on Decision and Control, Phoenix, 1999
- Schulmeister, Rolf: Virtuelle Universität - Virtuelles Lernen, Oldenbourg, München, 2001
- Schulmeister, Rolf: Grundlagen hypermedialer Lernsysteme, Oldenbourg, München, 2002

- Schulmeister, Rolf: Lernplattformen für das virtuelle Lernen - Evaluation und Didaktik, Oldenbourg, München, 2003
- Schulmeister, Rolf: "Diversität von Studierenden und die Konsequenzen für E-Learning", in: Campus 2004 - Kommen die digitalen Medien an den Hochschulen in die Jahre (Hrsg: Carstensen, Doris, Barrios, Beate), Waxmann, Münster, 2004
- Schwarzenberger, Astrid: "Studiendauer in zweistufigen Studiengängen - Ergebnisse eines internationalen Vergleichs", in: Der Bologna-Prozess im Spiegel der HIS-Hochschulforschung (Hrsg: Leszczensky, Michael; Wolter, Andrä), Eigenverlag, Hannover, 2005
- Sesink, Werner: "Pädagogische Kriterien für die Qualität von E-Learning-Systemen", in: Neue Medien in der Aus- und Weiterbildung von Bauingenieuren und Architekten (Tagungsband) (Hrsg: Grübl, Peter; Klaus, Hans G.), Eigenverlag, Darmstadt, 2003
- Steinborn, Thomas: "Analyse des universitären Lern- und Arbeitsverhaltens von Studierenden des Bauingenieurwesens", Dissertation, Technische Universität Darmstadt, Fachbereich Bauingenieurwesen und Geodäsie, Darmstadt, 2007
- Steinborn, Thomas; Lange, Jörg: "'Produktionsverfahren im Stahlbau' - online - Lehrveranstaltung via Internet - ein Versuch", in: Bautechnik, Band 84, 2007, S. 254-258
- Steinebach, Christoph: "Neue Theorie - neue Praxis? Perspektiven der Hochschuldidaktik aus psychologischer Sicht", in: Psychologie Lehren und Lernen. Beiträge zur Hochschuldidaktik (Hrsg: Steinebach, Christoph), Universitätsverlag Winter, Heidelberg, 2005
- Stojanovic, Ljiljana; Staab, Steffen; Studer, Rudi: "eLearning based on the Semantic Web", in: Proc. of WebNet2001 - World Conference on the WWW and Internet, Orlando, 2001
- Traeger, Dirk H.: Einführung in die Fuzzy-Logik, B.G. Teubner, Stuttgart, 1994
- Tribelhorn, Thomas: Crashkurs e-learning - Lerninhalte (2008), <http://www.crashkurs-elearning.ch/html/lerninhalte.htm> (nachgewiesen am 23.06.2008)
- Trommer, Jochen: "Morphologie", in: Computerlinguistik und Sprachtechnologie - Eine Einführung (Hrsg: Carstensen, K.-U.; Ebert, Ch.; Endriss, C.; Jekat, S.; Klabunde, R.; Langer, H.), Spektrum Akademischer Verlag, Heidelberg, 2001
- TU Darmstadt: Vorlesungsverzeichnis (SS 2008) (2008), <https://qis.pvw.tu-darmstadt.de/qisserver/rds?state=wtree&search=1&category=veranstaltung.browse&menuid=lectureindex> (nachgewiesen am 18.06.2008)
- Weber, Hubert: Einführung in die Wahrscheinlichkeitsrechnung und Statistik für Ingenieure, B. G. Teubner, Stuttgart, 1992

Wehling, Martin: "Multimediale Netzwerke im Bauwesen - Anwendung, Akzeptanz und Übertragbarkeit", Dissertation, Universität Duisburg-Essen, Institut für Bauphysik und Materialwissenschaft, Essen, 2008

Whitmire, Ethelene: "The relationship between undergraduates' epistemological beliefs, reflective judgment, and their information-seeking behavior", in: Information Processing and Management, Band 40, 2004, S. 97-111

Wikimedia Foundation Inc.: Neue Medien (2008), http://de.wikipedia.org/w/index.php?title=Neue_Medien&oldid=47007386 (nachgewiesen am 19.06.2008)

Wörner, Alexander: Lehren an der Hochschule - Eine praxisbezogene Anleitung, VS Verlag für Sozialwissenschaften, Wiesbaden, 2006

Anhang A: Verzeichnis verwendeter Formelzeichen

Formelzeichen	Bedeutung
a	Anzahl der Sätze bzw. Paarungen benachbarter Worte, die Term A enthalten
$A_{\text{Anim.}}$	auf die Textlänge bezogener Anteil von Animationen in einem Cluster
A_{Bild}	auf die Textlänge bezogener Anteil von Abbildungen in einem Cluster
$A_{\text{Det.}}$	Anteil der Determinativkomposita am Gesamttext
$A_{\text{FT}}(d_i)$	Anteil der ursprünglichen Suchbegriffe, der im Text des Dokuments d_i vorkommt
A_i	Teilfläche
$a_{m,i}$	mittlerer Abstand einer Wissensseinheit i zu den übrigen Wissensseinheiten in einem Cluster
$a_{m,\min}$	minimaler Abstand einer Wissensseinheit i zu den übrigen Wissensseinheiten in einem Cluster
$A_{\text{selt.}}$	auf die Textlänge bezogener Anteil seltener Begriffe in einem Cluster
$A_{\text{spez.}}$	Anteil der Begriffe mit wenigen Kookkurrenten
$A_{\text{UE}}(d_i)$	Anteil der ursprünglichen Suchbegriffe, der in der Überschrift des Dokuments d_i vorkommt
$A_{<100,i}$	Anteil von Begriffen mit weniger als 100 Kookkurrenten am Gesamttextumfang der Wissensseinheit i
$A_{<100,\min}$	minimaler Anteil von Begriffen mit weniger als 100 Kookkurrenten am Gesamttextumfang einer Wissensseinheit in einem Cluster
b	Anzahl der Sätze bzw. Paarungen benachbarter Worte, die Term B enthalten
a	Parameter
b	Parameter
b	empirischer Regressionskoeffizient
C	Parameter für ein Konfidenzintervall
$\cos(D_i, D_j)$	Kosinus des Winkels zwischen den Dokumentvektoren D_i und D_j
d_i	Dokument i eines Textkorpus
D	Parameter für ein Prognoseintervall
D_i	Dokumentvektor zum Dokument d_i
E_m	Mittelwert des mittleren Bereichs der Eingangsgröße E
E_{oben}	Übergang in den konstanten Abschnitt des hohen Bereichs der Eingangsgröße E
E_{unten}	Übergang vom konstanten Abschnitt des geringen Bereichs der Eingangsgröße E
E_0	Ergebnismenge nach der Suche
E_1	Ergebnismenge vor der Gliederung
E_2	Ergebnismenge des fertigen Lehrpfads
$f_{i,m}$	Frequenz des Terms t_i im Dokument d_m
FV	Dokumentvektor der Kookkurrenzen der Suchbegriffe, mit denen eine Wissensseinheit gefunden wurde
G_C	Grenzwert für die Zugehörigkeit zu einem Cluster
$G_{C,n}$	Grenzwert für die Zugehörigkeit zu einem Cluster vom Umfang n
h	Häufigkeit

Formelzeichen	Bedeutung
H_i	Hierarchiewert
IDF	Inverse Dokumentfrequenz
k	Anzahl der Sätze bzw. Paarungen benachbarter Worte, die die Terme A und B enthalten
k	Anzahl der Unterthemen
K	mittlere Kookkurrentenanzahl
L	Sollwert der Lehrpfadlänge
l_T	durchschnittliche Textlänge
MW	Mittelwert
$MW_{sig(A,B),NB}$	Mittelwert der Signifikanzen der Nachbarkookkurenzen
$MW_{sig(A,B),S}$	Mittelwert der Signifikanzen der Satzkookkurenzen
n	Anzahl aller Sätze bzw. Paarungen benachbarter Worte des betrachteten Textkorpus
n	Stichprobenumfang
N	Umfang eines Textkorpus
$nf_{i,m}$	normalisierte Termfrequenz
OV	Dokumentvektor der ursprünglichen Suchbegriffe inkl. Kookkurrenten
P	Precision
$P_{0,unten}$	unterer Grenzwert des Prognosestreifens an der Stelle VS_0
P'	Precision unter Einbeziehung lernzielähnlicher Wissenseinheiten
r	empirischer Korrelationskoeffizient
R	Relevanz einer Wissenseinheit für einen Lehrpfad
R_n	Rangplatz einer Wissenseinheit hinsichtlich des Unterthemas n
s	Standardabweichung
S	Signifikanzparameter
$sig(A,B)$	Signifikanzmaß der Kookkurenz zweier Terme A und B
$sig(A,B)_{NB}$	Signifikanz der Nachbarkookkurenzen
$sig(A,B)'_{NB}$	Signifikanz der Nachbarkookkurenzen, angepasst an die Größenordnung der Satzkookkurenzen
$sim_{di}(d_i)$	Ähnlichkeit zwischen zwei Dokumentvektoren D_i und D_j
$sim_{dW}(d_{LP})$	Ähnlichkeit der Seite d_{LP} in einem Lehrpfad mit einer (potentiellen) weiterführenden Seite d_W
$sim_{FV}(d_i)$	Ähnlichkeit des Dokuments d_i mit dem Vektor der Kookkurrenten der Suchbegriffe, mit denen das Dokument d_i gefunden wurde
$sim_{kor}(d_i)$	korrigierte Ähnlichkeit des Dokuments d_i zu einem Vektor
$sim_{max}(d_i)$	maximale Ähnlichkeit des Dokuments d_i zu einem Vektor
$sim_{OV}(d_i)$	Ähnlichkeit des Dokuments d_i mit dem Vektor der ursprünglichen Suchbegriffen OV
$sim_{SV}(d_i)$	Ähnlichkeit des Dokuments d_i mit dem erweiterten Suchvektor SV
$sim_{UT}(d_i)$	mittlere Ähnlichkeit des Dokuments d_i mit allen Unterthemen
$sim_{UT,n}(d_i)$	Ähnlichkeit des Dokuments d_i mit dem Unterthema n
$sim_{vorh}(d_i)$	tatsächlich berechnete Ähnlichkeit des Dokuments d_i zu einem Vektor

Formelzeichen	Bedeutung
$\text{sim}_{xv}(d_i)$	Ähnlichkeit des Dokuments d_i mit einem der o.a. Vektoren
SV	erweiterter Suchvektor
t	Größe des Vokabulars eines Textkorpus
t_i	Term i
$t_{n-2;1-\alpha/2}$	Quantil der t-Verteilung mit $n - 2$ Freiheitsgraden für das zweiseitige Signifikanzniveau α
$u_{0,95}$	95%-Quantil der Standardnormalverteilung
VS	Verknüpfungsstärke
w	Transformation, $w = \ln(VS)$
$w_{i,m}$	Gewichtung eines Terms t_i in einem Dokument d_m
$W_{99,5}$	Wert, unterhalb dessen 99,5% der Ähnlichkeiten aller gefundenen Dokumente zu einem Vektor (FV, OV, SV) liegen
x	ab/n
x_i	x-Koordinaten der Schwerpunkte der Teilflächen bei der Defuzzifizierung
x_R	Gesamtschwerpunkte der Teilflächen bei der Defuzzifizierung
y	Transformation, $y = P$ bzw. $y = P'$
$1 - \cos(\alpha)$	Clusterdifferenzmaß mit dem Winkel α
ΔL	maximale Anzahl von Zusatzseiten zu Unterthemen
Δ_{SPV}	Clusterdifferenzmaß einer Wissenseinheit zum Schwerpunktsvektor SPV
μ	Mittelwert der Grundgesamtheit
$\mu_A(y)$	Zugehörigkeit eines Elements y zu einer unscharfen Menge A
ρ	Korrelationskoeffizient für die Grundgesamtheit
σ^2	Varianz der Grundgesamtheit

Anhang B: Glossar wichtiger Fachbegriffe

Im Folgenden sollen einige der in dieser Arbeit auftretenden Fachbegriffe, die keinen direkten Bezug zum Bauwesen haben, erläutert werden.

Kompositum	Ein Kompositum bezeichnet ein aus zwei oder mehr einzelnen Worten, den Konstituenten, zusammengesetztes Wort. Man unterscheidet im Wesentlichen Determinativkomposita, bei denen einer der Konstituenten durch den anderen näher bestimmt wird, sowie Kopulativkomposita, bei denen beide Konstituenten gleichberechtigt sind.
Kookkurenz	Eine Kookkurenz bezeichnet das gemeinsame Auftreten zweier Worte in einem definierten Textabschnitt. Im Rahmen der vorliegenden Arbeit sind das entweder benachbarte Worte oder Worte, die in demselben Satz auftreten.
Lehrpfad	Ein Lehrpfad bezeichnet eine Zusammenstellung von → Wissensseinheiten zu einer Lerneinheit, die von Lernern bearbeitet werden kann. Der Lehrpfad ist hierarchisch strukturiert. In der Regel sollte ein Lehrpfad sequentiell bearbeitet werden.
Inhaltsbasis	Im Rahmen der vorliegenden Arbeit umfasst eine Inhaltsbasis die in einem E-Learning-System zu einem Wissensgebiet enthaltenen Dokumente und die daraus mit textstatistischen und computerlinguistischen Methoden gewonnenen Informationen. Der Begriff ist angelehnt an die „Wissensbasis“, einen Fachausdruck aus dem Wissensmanagement. Eine Wissensbasis umfasst aber neben Inhalten auch explizite Regeln zu deren Verknüpfung sowie, je nach Definition, auch noch das Programm, um die Inhalte und Regeln zu nutzen.
Precision	Anteil der für eine Suchanfrage relevanten Seiten an den Suchergebnissen
Recall	Anteil aller relevanten Seiten in der → Inhaltsbasis an den Suchergebnissen
Wissensdomäne	Eine Wissensdomäne bezeichnet das zu einem bestimmten Fachgebiet vorhandene Wissen. Je nach Kontext kann damit das in einem E-Learning-System enthaltene Wissen oder das gesamte Wissen des Fachgebiets gemeint sein. Statt Wissensdomäne kann auch der Begriff „Wissensgebiet“ verwendet werden.
Wissenseinheit	Eine Wissensseinheit ist nach dem Konzept des WiBA-Net die kleinste für die Lernenden verfügbare Einheit. Wissensseinheiten sind HTML-Seiten. Sie können Fließtext und/oder multimediale Elemente enthalten. Jede Wissensseinheit hat zudem eine Überschrift. In einer Wissensseinheit wird immer nur ein Inhalt behandelt. Wissensseinheiten sind kontextfrei, können also prinzipiell frei kombiniert werden.

Anhang C: Wortendungen

Beim Stemming der in den Wissensseinheiten enthaltenen Worte wurden die folgenden Wortendungen berücksichtigt:

- heiten
- keiten
- endem
- enden
- ungen
- heit
- keit
- nden
- den
- dem
- ung
- de
- em
- en
- er
- es
- se
- e
- n
- r
- s
- t

Anhang D: Ausschlussworte

Die nachfolgende Liste stellt die fünfzig häufigsten Worte im Textkorpus des WiBA-Net dar. Die Begriffe „Stahl“, „Beton“, „m“, „Holz“, „Prozent“ und „Beispiel“ werden dabei übersprungen.

Tabelle D.1: Zusammenstellung der häufigsten Begriffe im WiBA-Net (Exemplarische Ergänzungen in eckigen Klammern bei unklaren Wortstämmen)

Rang	Anzahl	Begriff
1	7902	die
2	7591	der
3	6203	und
4	3383	in
5	3158	von
6	2400	des
7	2385	werden
8	2184	das
9	2125	mit
10	2075	ist
11	1874	zu
12	1870	bei
13	1859	im
14	1807	den
15	1742	fuer
16	1712	durch
17	1589	wird
18	1553	eine
19	1435	oder
20	1391	sind
21	1304	aus
22	1295	auf
23	1280	sich
24	1213	als
25	1205	ein
26	1155	zum
27	969	einer
28	941	dem

Rang	Anzahl	Begriff
29	920	nach
30	811	es
	(786)	(beispiel)
31	784	auch
32	770	zur
33	733	an
34	726	sie
	(693)	(stahl)
35	687	nicht
36	656	kann
	(654)	(beton)
37	652	koenn[en]
38	643	sei[n]
39	613	bis
	(588)	(m)
40	584	unt[en]
41	577	einem
42	553	gros[s]
43	511	wie
	(502)	(prozent)
44	499	eines
45	481	diese
46	470	um
47	470	man
48	468	hoh[er]
49	459	so
	(452)	(holz)
50	446	dass

Anhang E: Im Projekt WiBA-Net integrierte Hochschullehrer

Projektleitung:	Prof. Grübl Institut für Massivbau, TU Darmstadt
Inhalte:	Prof. Grübl Institut für Massivbau, TU Darmstadt Prof. Setzer Institut für Bauphysik und Materialwissenschaft, Universität Duisburg-Essen Prof. Reinhardt Institut für Werkstoffe im Bauwesen, Universität Stuttgart Prof. Franke Institut für Baustoffe, Bauphysik und Bauchemie, TU Hamburg-Harburg Prof. König Institut für Massivbau und Baustofftechnologie, Universität Leipzig Prof. Hillemeier Fachgebiet Baustoffe und Baustoffprüfung, TU Berlin Prof. Schäfer Institut für Massivbau, TU Darmstadt Prof. Pahl Institut für Grundlagen des Bauens und Planungsmanagement, Universität Leipzig Prof. Vormwald Institut für Stahlbau und Werkstoffmechanik, TU Darmstadt Prof. Seeger Institut für Stahlbau und Werkstoffmechanik, TU Darmstadt
Technik:	Prof. Encarnaç�o Fraunhofer IGD, Darmstadt Prof. Mühlhäuser Fachgebiet Telekooperation, TU Darmstadt
Pädagogik:	Prof. Sesink Institut für Allgemeine Pädagogik und Berufspädagogik, TU Darmstadt

Anhang F: Liste der Wissenseinheiten des WiBA-Net

Nachfolgend findet sich eine um doppelte Einträge und Lehrzielseiten bereinigte Aufstellung der im WiBA-Net enthaltenen Wissenseinheiten. Diese bilden die inhaltliche Basis für die linguistischen und textstatistischen Verfahren der vorliegenden Arbeit.

"House of the Future"	Allgemeines zu Abkühlkurven	Allgemeines zur Plastizität von Vielkristallen
"textile-block"-System von Frank Lloyd Wright	Allgemeines zu Arten und Beschreibung von Kristallfehlern	Allgemeines zur Spannung
(1) Heizelementschweißen	Allgemeines zu Betonstahlmatten	Allgemeines zur Wärmebehandlung
(1) Verbindungsarten einer Holzkonstruktion	Allgemeines zu Edelstahl	Alterung von Beton
(1) Vibrationsschweißen	Allgemeines zu Flächenfehlern	Aluminium
(2) Holz-Anschlussknoten – Verkämmungen	Allgemeines zu Gitterparameter	Aluminium-Legierungen
(2) Holz-Anschlussknoten – Versätze	Allgemeines zu Gitterparametern	Aluminiumprodukte im Bauwesen
(2) Holz-Anschlussknoten – Zapfenausbildung	Allgemeines zu Gleitsystemen	Andere legierte Stähle
(2) Laserschweißen	Allgemeines zu Korngrenzen	Anforderungen an Konstruktionen
(2) Ultraschallschweißen	Allgemeines zu Kristalliten	Anforderungen an den Brandschutz
(3) Hölzerne Verbindungsmittel	Allgemeines zu Linienfehlern	Anforderungen an den Feuchteschutz
(3) Warmgas- und Extrusionsschweißen	Allgemeines zu Millerschen Indizes	Anforderungen an den Schallschutz
AV-T-Kurve / Bereichsdefinitionen	Allgemeines zu Millerschen Indizes für Gitterebenen	Anforderungen an den Wärmeschutz
Abbinden mit chemischer Reaktion – Reaktionskleber und ihre Umgebungsbedingungen	Allgemeines zu Millerschen Indizes für Richtungen einer Ebene	Anforderungen an den Werkstoff
Abbindevorgänge beim Kleben	Allgemeines zu Phasendiagrammen	Anforderungen an die Zusammensetzung von "Walz"-Asphalt
Abbrennstumpfschweißen/Gaspressschweißen	Allgemeines zu Proportionalstäben	Angeregte Atomzustände / Promotion
abfließende Hydratationswärme	Allgemeines zu Punktfehlern	Angriff durch nichtmetallisch-anorganische Baustoffe
Abgrenzung zur Ermüdungsmechanik	Allgemeines zu Schlißbildern	Anhydritestrich
Abhängigkeit der Festigkeit von der Rohdichte	Allgemeines zu Spannstählen	Animation eines Zugversuchs
Abschreckalterung	Allgemeines zu Stapelfehlern	Animation zum Hebelgesetz
Abschrecken durch Löschmittel	Allgemeines zu Verbindungsarten	Animation zur Bestimmung des Gehalts an fester Phase (Mischkristall)
Abschreckmittel	Allgemeines zu Verbindungsmittel (VBM)	Animation zur Bestimmung des Gehalts an flüssiger Phase (Schmelze)
Absolute und relative Geschwindigkeit	Allgemeines zu grauem Gusseisen	Animation zur Entwicklung von Phasendiagrammen
Abstandsfaktor	Allgemeines zu weißem Gusseisen – Temperguss	Anlassen
Abstoßungskraft in der ionischen Bindung	Allgemeines zum Begriff der Druckfestigkeit	Anlaßsprödigkeit
Adhäsion und Kohäsion	Allgemeines zum Betonstahl	Anodische Fremdmetallüberzüge
Ästigkeit und andere Holzfehler	Allgemeines zum Bezeichnungssystem	Anodische Teilreaktion
Akkustische Eigenschaften von Glas	Allgemeines zum Bezeichnungssystem für Betonstahl	Anodische und kathodische Teilreaktion
Alkali-Kieselsäure-Reaktion	Allgemeines zum Eisen-Kohlenstoff-Diagramm	Ansteifen, Erstarren & Erhärten
Allgemeine Baustähle	Allgemeines zum Gleiten	Anstriche
Allgemeine Formgebungsverfahren – Urformen	Allgemeines zum Gusseisen	antiklastische Formen
Allgemeine Lösung	Allgemeines zum Periodensystem der Elemente (PSE)	Anwendung von Spannstahl
Allgemeine Ursachen der Metallkorrosion	Allgemeines zum Schmid-Faktor	Anwendungen von Hochleistungsbaustoffen
Allgemeines	Allgemeines zum Wöhlerversuch	Anwendungsbeispiel – Blockhausbau
Allgemeines – Feste Aufkohlungsmittel	Allgemeines zum Zustandsfeld	Anwendungsbeispiel – Hallenbau mit Brettschichtholz (BSH)
Allgemeines Verformungsverhalten	Allgemeines zur Dehnung	Anwendungsbeispiel – Holzfassaden
	Allgemeines zur Ermüdungsfestigkeit	Anwendungsbeispiel – Holzskelettbau
	Allgemeines zur Metallbindung	Anwendungsbeispiel – Holztafelbau

Anwendungsbeispiel – Verbindungsmittel	Auslagerungstemperatur	Beispiel: Bestimmung der Millersche Indizes für Ebenen
Anwendungsbeispiele für den Werkstoff Holz im Bauwesen	Ausscheidung	Beispiel: Bestimmung der Millersche Indizes für Richtungen einer Ebene
Anwendungsgebiete von Naturstein	Außenrüttler	Beispielaufgabe zur Hydratationsgradberechnung
Anzahl der Atome in einer Elementarzelle	Außergewöhnliche Lasten	Beispielaufgabe zur Nernst-Gleichung
Anzahl der bekannten Elemente und ihrer Ordnungszahlen	Austenit	Beispiele – Bezeichnung mit Kurznamen nach Verwendung
Anziehungskraft der ionischen Bindung	Auswahl geeigneter Werkstoffe	Beispiele für die Auswirkung der Hundschen Regel
apparative Sanierungsmaßnahmen	Auswahl von Holzprodukten	Beispiele zum Bezeichnungssystem von Betonstahlmatten
Arbeitsvermögen	Auswirkungen unterschiedlicher E-Moduli	Beispiele zur Auswirkung der Hundschen Regel II
Arten und Erscheinungen der Metallkorrosion	Balkenbrücken	Beizen
Asphalt – Ausgangsstoffe für die Mischung	Bank of China	Bekämpfung und Sanierung befallener Holzbauteile
Asphalt – Besondere Eigenschaften	Basiseinheiten und abgeleitete Größen	Belegung der Schalen bei Haupt- und Nebengruppenelementen
Asphalt mit Bindemittelüberschuß – Gußasphalt	Baugrundversagen	Belüftungselement
Asphaltmischungen mit Ausfallkörnung	Baulicher Holzschutz	Berechnung der Frischbetontemperatur
Atombindung – typische Vertreter	Baumstützen	Berechnung des Spannungsintensitätsfaktors
Atommodell nach Nils Bohr	Bauphysik	Berechnung des pH-Wertes
Aufbau des Hochofens	Baustähle und Werkzeugstähle	Berechnung von Membrankonstruktionen
Aufbau einer Bitumenbahn	Baustoffklassen	Berechnungsgrundlagen
Aufbau und allgemeine Eigenschaften von Kunststoffen	Bauteilbeanspruchungen und Grenzzustände	Beschichtung von Geweben
Aufbau von Holz – Elementzusammensetzung	Beanspruchung in Stablängsrichtung (Richtung der wirkenden Normalkraft)	Beschichtung von Glas
Aufbau von Holz – Fette und Harze	Beanspruchungsarten	Beschichtungen für Textilien und Membranen
Aufbau von Holz – Hemizellulose	Bearbeitbarkeit	Beschichtungsaufbau
Aufbau von Holz – In Stammlängsrichtung orientierte Röhrenstruktur	Bearbeitung von Glas	Beschränkte Löslichkeit
Aufbau von Holz – Lignin	Beginn der Eisengewinnung – Der Renofen	Besondere Verbindungsmittel
Aufbau von Holz – Zellulose	Begrenzung der Rissbreiten	Besonderheit für kubische Kristalle
Aufbauprinzip	Begrenzung der Schwingungen	Besonderheiten von Hochhäusern
Aufbereitung der Gleichgewichtsgleichungen zur Ermittlung geschlossener Lösungen	Begrenzung der Verformungen	Besonderheiten von Zementestrich
Aufbereitung der Rohstoffe	Begriff der Säure und Base	Beständigkeit
Aufgaben und Arten des Korrosionsschutzes	Begriff der starken und schwachen Base	Bestimmung der Lichtgeschwindigkeit nach Bradley
Aufgespritzte Überzüge	Begriff der starken und schwachen Säure	Bestimmung des Ausbreitmaßes
Aufhärtung	Begriffsbestimmung – Zement	Bestimmung des Rauminhalts bei gleichmäßig geformten Körpern
Aufkohlungsmittel	Behaglichkeit im Bauwerk	Bestimmung des Rauminhalts bei ungleichmäßig geformten Körpern
Aufschweißbiegeversuch	Behaglichkeit in Wohnräumen	Bestimmung des Slumpmaßes
Auftragsschweißen & Verbindungsschweißen	Beispiel Smithdiagramm	Bestimmung des Verdichtungsmaßes
Auftrieb	Beispiel Studentenwohnheim Karlshof	Bestimmung des Vébémaßes
Ausbreitmaßklassen	Beispiel Wöhlerdiagramm	Bestimmung des Zustandsdiagramms aus den Abkühlkurven
Ausbreitversuch	Beispiel für die Belegung der Orbitale	Beton
Ausgangsmaterial	Beispiel für die Berechnung einer Rohstoffmischung	Betondruckfestigkeit / Prüfkörpergestalt und -abmessungen
Ausgangspunkt des Periodensystems	Bsp. für die Bezeichnung von Betonstabstahl	Betondruckfestigkeitsprüfung
Ausgangsstoffe	Beispiel für die Bezeichnung von Bewehrungsdraht	Betondruckfestigkeitsprüfung / Filme
Ausgangsstoffe Glas	Bsp. für die unterschiedlichen Dichtebegriffe	Betoneinbau
Ausgangsstoffe des Betons (Zementbeton)	Beispiel sichtbare Kristallstruktur (Zink)	Betonfestigkeitsklassen
Ausgewählte Kunststoffe und deren Eigenschaften	Beispiel zu Momentan- und Durchschnittsgeschwindigkeit	
	Beispiel zur Absolut- und Relativgeschwindigkeit	
	Bsp. zur Bestimmung der Anzahl der Orbitale	
	Beispiel zur Gibbssche Phasenregel	

Betonstähle	Biegeeweiche Membranecken	Carbonatisierung – Überblick
Betonstahlverbindungen	Biegezugfestigkeit	Carbonatisierung / Einfluss regelmäßiger Befeuchtung
Betonzusatzmittel / Beschleuniger (BE)	Biegung	Carbonatisierung / Einflussfaktoren
Betonzusatzmittel / Betonverflüssiger (BV)	Bindemittel	Carbonatisierung / Einführung
Betonzusatzmittel / Chromatreduzierer (CR)	Bindigkeit	Carbonatisierung / Experimentelle Bestimmung der Tiefe 1
Betonzusatzmittel / Dichtungsmittel (DM)	Bingham-Hook-Körper	Carbonatisierung / Experimentelle Bestimmung der Tiefe 2
Betonzusatzmittel / Dosierung	Bingham-Körper	Carbonatisierung / Experimentelle Bestimmung der Tiefe 3
Betonzusatzmittel / Einleitung	Bitumen – Anwendungen	Carbonatisierung / Experimentelle Bestimmung der Tiefe 4
Betonzusatzmittel / Einpresshilfen (EH)	Bitumen – Ein Erdölprodukt	Carbonatisierung / Experimentelle Bestimmung der Tiefe 5
Betonzusatzmittel / Fließmittel (FM)	Bitumen – Herstellung	Carbonatisierung / Geschwindigkeit
Betonzusatzmittel / Luftporenbildner (LP)	Bitumen – Verarbeitungsformen	Carbonatisierung / Reaktionsgleichung
Betonzusatzmittel / Recyclinghilfen für Waschwasser (RH)	Bituminöse Abdichtungen	Carbonatisierung / Schadensvermeidung
Betonzusatzmittel / Stabilisierer (ST)	Bituminöse Überzüge	Carbonatisierung / Witterungseinfluss
Betonzusatzmittel / Verzögerer (VZ)	Blasstahlverfahren (Windfrischverfahren)	Carbonatisierung / Zusammenhang zwischen Carbonatisierung und Trocknung
Betonzusatzmittel / Wirkungsgruppen	Bleiprodukte im Bauwesen	Carbonitrierhärten
Betonzusatzstoffe / Einführung	Blockguss	Charakterisierung der Poren über deren Größe
Betonzusatzstoffe / Inerte Stoffe und Pigmente	Blockhausbau	Charakteristische Eigenschaften – Anisotropie
Betonzusatzstoffe / Kennwerte	Blockieren der Gleitebenen	Charakteristische Eigenschaften – Feuchteverhalten
Betonzusatzstoffe / Latent-hydraulische Stoffe	Böschungs- und Geländebruch	Charakteristische Eigenschaften – Schwinden und Quellen
Betonzusatzstoffe / Organische Stoffe	Bogenbrücken	Charakteristische Eigenschaften von Stoffen mit Metallbindung
Betonzusatzstoffe / Puzzolanische Stoffe	Bohrung von Glas	Chemisch härtende Klebstoffe
Betonzusatzstoffe / Silicastaub	Boudouard-Gleichgewicht	Chemische Angriff auf bewehrte Betonbauteile
Betonzusatzstoffe / Steinkohlenflugasche	Brandklassen	Chemische Beständigkeit von Glas
Betonzusatzstoffe / Wirkungsgruppen	Brandschutzmaßnahmen	Chloridkorrosion
Betonzusatzstoffe / Zusatzstoffarten nach DIN EN206-1/DIN 1045-2	Brandverhalten von Holz	Compliance-Methode
Betonzusatzstoffe / anorganisch-mineralische Zusatzstoffe	Bravais-Gitter	Dachdämmung
Betonzusatzstoffe / inaktive Stoffe	Brechpunkt	Dämmung der Kellerdecke
Betonzuschlag / Einführung	Brechung des Lichts	Darstellung der d-Orbitale
Betrachtung nicht symmetrischer Fälle	Brechung des Lichts II	Darstellung der p-Orbitale
Betrachtung symmetrischer Fälle	Brechungskonstante	Darstellung der s-Orbitale
Beurteilen und Bewerten von Holzschutzmitteln	Brennen des Zements	Darstellung des viskosen Fließens
Beweglichkeit der Ionen	Brennhärten	Das Ampere
Bezeichnen von Bitumenbahnen	Brennstoffe	Das Bessemer-Verfahren von 1855
Bezeichnung mit Kurznamen nach Verwendung	Bruch infolge Ermüdung	Das Einheitensystem SI
Bezeichnung mit Kurznamen nach chemischer Zusammensetzung	Bruchdehnung	Das Elastische Verhalten
Bezeichnung nach Werkstoffnummern	Bruchdehnung (A)	Das Kelvin
Bezeichnung nichtrostender Stähle	Bruchdehnung (A) beim kaltverformten Stahl	Das Kilogramm
Bezeichnung von NE-Metallen	Bruchdehnung (A) beim warmverformten Stahl	Das Maxwell Modell
Bezeichnungsbeispiele für Edelstähle	Brucheinschnürung (Z)	Das Meter
Bezeichnungssystem von Betonstahlmatten	Brucheinschnürung (Z) für einen kaltverformten Stahl	Das Mol
Biegefähigkeit und Rückbiegeversuch	Brucheinschnürung (Z) für einen warmverformten Stahl	Das Pascal
Biegen von Betonstählen	Brücken	Das Pirelli Hochhaus
biegesteife Membranecken	Buche – Eigenschaften – Beständigkeit – Verwendung	Das Plastische Verhalten
	Burgersvektor	
	CFK – Anwendungen carbonfaserverstärkter Kunststoffe	
	CFK – Anwendungsbeispiel carbonfaserverstärkter Kunststoffe	
	CFK – Eigenschaften carbonfaserverstärkter Kunststoffe	
	CN Tower (553 m)	
	CO2-Reduktion im Wohnungsbau	

Das Pourbaix-Diagramm für Eisen (Herleitung)	Der Einfluss des w/z-Werts auf die Betondruckfestigkeit	Die technische Festlegung des Amperes
Das Puddelverfahren seit 1783	Der Fachwerkbau – Allgemeines	Diffusion
Das Siemens-Martin-Verfahren	Der Fachwerkbau – Konstruktion	Diffusionsglühen
Das Smith Diagramm	Der Fachwerkbau – Schwächen der Konstruktion	Diffusionsverfahren
Das Verfahren der Zementherstellung	Der Fernsehturm in Moskau-Ostankino (537 m)	Dipol
Das Viskose Verhalten	Der Hausbock	Direkte Lasteinwirkungen
Das Voigt-Kelvin Modell	Der Holzmassivbau	Direktreduktion
Das Windfrischverfahren	Der Holzmassivbau – Blocktafelbauweise	Dissoziation
Das Wöhlerdiagramm – Versuchsablauf	Der Holzmassivbau – Brettstapelbauweise	Doppelter Stapelfehler
Das erste Kunststoffrezept – Kunsthorn	Der Holzmassivbau – Dickholz-System	Drahtseile
Das ideale Verformungsverhalten	Der Holzmassivbau – Wand-Baukastensystem	Druckfestigkeit
Das optische Spektrum	Der Holzrahmenbau	Druckfestigkeit der reinen Klinkerphasen
Dauer der Konsistenzbeibehaltung	Der Holzrahmenbau und Holztafelbau – Allgemeines	Druckfestigkeitsbestimmung / Prüfalter
Dauerfestigkeit und Grenzlastspielzahl	Der Holzskelettbau – Allgemeines	Druckfestigkeitsklassen für Leichtbeton
Dauerhaftigkeit	Der Holzskelettbau – Konstruktion	Druckfestigkeitsklassen für Normal- und Schwerbeton
Dauerhaftigkeit in der Praxis – kombinierte Beanspruchungen	Der Holzskelettbau – Zangen, Pfosten und Riegel	Dübel
Dauerhaftigkeit von bewehrten Betonkonstruktionen	Der Holztafelbau	Düsseldorfer Brückenfamilie von Fritz Leonhardt
Dauerschwingbeanspruchung	Der Mensch und das Holz	Duktiles Bauteilversagen
Dauerschwingfestigkeit (Dauerfestigkeit)	Der Newtonsche Körper	Duktiles Versagen von Stahlbetonbauteilen mit normalem Bewehrungsgehalt
Dauerschwingversuch	Der Prozess der Formfindung	Duplexsysteme
Dauerstandverhalten und Dauerschwingfestigkeit	Der Sendeturm von Schabolvka	Durchstrahlverfahren
Definition Dehnungen	Der Werkstoff Holz im Bauwesen	Duroplaste
Definition Edelstahl	Der Wöhlerversuch	Dynamisch-Mechanische-Analyse (DMA)
Definition Komponenten	Der echte Hausschwamm	E-Modul von Natursteinen
Definition Kristall	Der moderne Glasbau	Edelgaskonfiguration in der Atombindung
Definition Phasen	Der moderne Mauerwerksbau	Edelstähle
Definition Spannungen	Desoxidation	Edelstahl im Bauwesen
Definition Wärmeausdehnung	Dichtest gepackte Ebenen	Eiche – Allgemeine Eigenschaften
Definition Wärmedurchlasswiderstand	Die Arten idealen Verformungsverhaltens	Eifelturm
Definition Wärmeleitfähigkeit	Die Beschleunigung a	Eigenschaften der Kunststoffe
Definition der Atombindung	Die Candela	Eigenschaften des Lichts
Definition der Ionenbindung	Die Dichte ρ	Eigenschaften und Anforderungen an Spannstähle
Definition der Porosität	Die Energie	Eigenschaften und Anwendungsmöglichkeiten bituminöser Baustoffe
Definition der Rein- und Rohdichte	Die Fläche A und das Volumen V	Eigenschaften von Aluminium
Definition des pH-Wertes	Die Geschwindigkeit v	Eigenschaften von Betonstahl
Definitionen	Die Grundbausteine der Modellrheologie	Eigenschaften von Edelstahl
Definitionen zu Betonstahl	Die Gusseisenperiode 1780-1850	Eigenschaften von Folien aus Kunststoff
Dehngrenze ($R_{p0,2}$)	Die Kraft	Eigenschaften von Gewebearten aus Kunststoff
Der Aufbau von Holz – Richtungsabhängigkeit	Die Leistung	Eigenschaften von Hüttensteinen
Der Aufbau von Holz – chemische Verbindungen	Die Schmiedeeisenphase- Erste Hängekonstruktionen	Eigenschaften von Kalksandsteinen
Der Baustoff Holz – seine charakteristischen Eigenschaften	Die Sekunde	Eigenschaften von Kupfer
Der Beginn der Stahlperiode	Die Stahlperiode- Neue Ideen, neue Dimensionen	Eigenschaften von Mauersteinen aus Normal- und Leichtbeton
Der Beginn des Mauerwerksbaus	Die Unbedenklichkeit für den Nutzer	Eigenschaften von Natursteinen
Der Begriff des Atoms	Die Verwendungsgeschichte von Steinen und Ziegeln im Mauerwerksbau	Eigenschaften von Porenbetonsteinen
Der Blockhausbau – Allgemeines		Eigenschaften von Zink
Der Blockhausbau – Konstruktion		Eigenstressungen

Eignung der Löschmittel	Einführung in die Geschichte des Mauerwerksbaus	Elastische Verformung beim warmverformten Stahl
Einachsige Beanspruchung	Einführung in die Geschichte von Eisen und Stahl im Bauwesen	Elastizität
Einbau des Betons	Einführung in die Maßeinheiten	Elastizität infolge Schubbelastung
Eindring- oder Penetrationsversuch	Einführung in die Thematik	Elastizitätsgrenze
Einfacher Stapelfehler	Einführung verschiedener Dichtebegriffe	Elastizitätsmodul
Einfamilienhaus (Baujahr 1950)	Einführung zu Kunststoffschweißen	Elastizitätsmodul (E-Modul)
Einflüsse auf das Elektrodenpotential	Einführung: Naturstein	Elastomer- und Plastomerbitumen
Einflüsse auf die Bruchzähigkeit	Einkomponentenklebstoffe	Elastomere
Einflüsse auf die Druckfestigkeit von Beton	Einlagerungsatom	Elektrode
Einflüsse auf die Konsistenz	Einleitung – Der Weg ins Kunststoffzeitalter	Elektrolyt
Einfluß der Atmosphäre auf die Korrosionsgeschwindigkeit	Einleitung – Geschichte der Metalle von der Steinzeit zur Neuzeit	Elektrolytische (galvanische) Überzüge
Einfluß der Holzfeuchte auf das Verformungsverhalten	Einleitung – Geschichte des Betonbaus	Elektronen- /Laserstrahlschweißen
Einfluss der Legierungselemente	Einleitung – Übersicht der Holzbauweisen	Elektronenaffinität
Einfluss der Luftfeuchtigkeit	Einleitung – Wahrnehmung-die plastische Qualität von Beton	Elektrostahlverfahren
Einfluss der Sauerstoffkonzentration im Elektrolyten	Einsatz von Kohlefasern im Brückenbau	Element
Einfluss der Temperatur	Einsatzgebiete von Aluminium	Elementarteilchen
Einfluss der Temperatur auf behandelten/kaltverformten Stahl	Einsatzgebiete von Kupfer	Elementarzelle
Einfluß der Temperatur auf das Verformungsmodul von Holz	Einsatzgebiete von Zink	Eloxalverfahren
Einfluss der Temperatur auf mechanische Eigenschaften	Einschnürungsbereich	Eloxieren (Anodisieren)
Einfluss der Witterung auf den Carbonatisierungsfortschritt	Einschnürungsbereich beim kaltverformten Stahl	Email
Einfluss des Verhältnis Anode/Kathode	Einteilung der Betone nach dem Erhärtungszustand	Empire State Building
Einfluss des pH-Werts	Einteilung der Kunststoffe	Endkriechzahl für Normalbeton und feuchte Umgebungsbedingungen
Einfluss von Alter und Lagerungsverhältnissen auf die Betondruckfestigkeit	Einteilung der Natursteine nach der Entstehung	Endkriechzahl für Normalbeton und trockene Umgebungsbedingungen
Einfluss von Querschnittsform und Werkstofftrennfestigkeit	Einteilung des Periodensystems in Blöcke	Endkriechzahl für feuchte Umgebungsbedingungen
Einflussgrößen	Einteilung des Periodensystems in Perioden	Endkriechzahl für trockene Umgebungsbedingungen
Einführung – Faserverbundwerkstoffe (FVV)	Einteilung in Festigkeitsbereiche	Energetische Betrachtung nach GRIF-FITH
Einführung – Formgebung und Fügen	Einteilung nach dem Beruhigungsgrad	Energiebeträge der Ionenbindung am Beispiel des Kaliumchlorid
Einführung – Handwerkliche Holzverbindungen	Einteilung nach dem Herstellungsprinzip	Energiefreisetzungsrate
Einführung – Ingenieurmäßige Verbindungsmittel	Einteilung nach der Konstruktion von Zellen aus Kunststoff	Entfetten
Einführung – Kleben mit Kunststoffen	Einteilung nach der Molekularstruktur	Entstehung von Kristalliten
Einführung – Konsistenz	Einteilung nach der Polarität	Entwicklung von Bauteilbeanspruchungen und Grenzzuständen
Einführung – Kunststoffe für Textilien und Membranen	Einteilung von Holzprodukten nach DIN 1052	Entwicklungsgeschichte des Glases
Einführung – Was sind Kunststoffe?	Einteilungen	Entwicklungsgeschichte des Mauerwerks vom 18. bis zum 20. Jhd.
Einführung Ästhetik und Gestaltung von Mauerwerk	Einteilungsprinzipien für Nichteisenmetalle	Entwicklungsgeschichte des Mauerwerksbaus
Einführung Grundlagen Estriche	Einwirkungen auf Konstruktionen	Entwicklungsgeschichte von Eisen und Stahl
Einführung Holz – Begriffe im Holzbau	Einzelbeispiele	Entwicklungsgeschichte von Textilien und Membranen
Einführung Konstruktion und Füge-techniken von Membranen	Eisdruck	Entwicklungslinie von Holzwerkstoffen
Einführung Raumzellen aus Kunststoff	Eisen und Stahl im Bauwesen	Entwicklungslinien des Holzskelettbaus
Einführung Textilien und Membranen	Elastische Verformung	Epoxidharze und MS-Polymere im Kühlcontainerbau
Einführung in das Themenfeld der abgeleiteten Maßeinheiten	Elastische Verformung beim kaltverformten Stahl	Erdbeben
Einführung in die Geschichte des Glasbaus		Erdölfraktionen aus der Destillation von Rohöl
		Ergebnisse Riedstadt
		Ergebnisse des Milikans-Versuchs
		Ergebnisse des Milikans-Versuchs II

Erhärten des Betons	Faserbeton / Glasfasern	Fördern des Betons
Erklärung der Kurzzeichen und Fachbegriffe (Holz)	Faserbeton / Herstellung	Folgen und Verfahren der Kaltverformung
Ermittlung der Querschnittsspannungen	Faserbeton / Kunststofffasern	Folien
Ermittlung des statischen E-Moduls	Faserbeton / Spannungs-Dehnungslinie	Form und Funktion – Tragwerke aus Beton (1)
Ermüdungsbruch	Faserbeton / Spritzverfahren	Form und Funktion – Tragwerke aus Beton (2)
Erneuerung der Heizungsanlage	Faserbeton / Stahlfaserarten	Formbarkeit von NE-Metallen
Erscheinungen beim Vergießen	Faserbeton / Stahlfasern	Formentwicklung von Membrantragwerken
Erstarrungsstörungen	Faserbeton / Tragverhalten	Formgebung – Handlaminieren und Ziehverfahren
Erze	Faserbeton / Verbundmechanismen	Formgebung von Stahl
Estrich auf Trennlagen	Faserbeton / Verhalten bei Druckbeanspruchung	Formteile
Estricharten	Faserbeton / Verhalten bei Explosions-, Schlag- und Stoßbeanspruchung	Frenkel-Paar
Estriche – Ausführung der Oberflächen	Faserbeton / Verhalten bei Zug- und Biegebeanspruchung	Frischbetoneigenschaften
Estriche – Materialien	Faserbeton / Wasserzementwert, Zementgehalt und Betonzusätze	Frischbetontemperatur
Estriche – Verbund zum Untergrund	Faserbeton / Zusammenwirken von Fasern und Matrix	Frischen
Estriche mit besonderen Funktionen	Faserbeton / Zuschlag	Frostdruck
Eutektikum (E)	Faserwerkstoffe	Fügenprinzipien von Membranen
Eutektoider Stahl	Fehler im Bohrschen Atommodell	Fügetechniken und Verarbeitung von Textilien und Membranen
Eutektoidischer Zerfall	Fenster	Fügetechniken von Textilien und Membranen
Exkurs Stahlbeton: Depassivierung	Fernmeldeturm in Hannover (133 m)	Funkmaste
Exkurs Stahlbeton: Karbonatisierung	Ferrit	Funktionalität
Exkurs Stahlbeton: Volumenvergrößerung	Festbeton	Futuro-Haus
Exkurs Stahlbeton: pH-Wert der Porenlösung	Festigkeit und Erhärtung des Zementsteins	GFK – Anwendungen glasfaserverstärkter Kunststoffe
Experimentelle Methode	Festigkeit, Dichte und Porosität von Natursteinen	GFK – Eigenschaften glasfaserverstärkter Kunststoffe
Expositionsklasse / X0	Festigkeitsabfall – Stahlverhalten bei Erwärmung	Gangart
Expositionsklassen / Einführung	Festigkeitskennwerte	Gasförmige Aufkohlungsmittel
Expositionsklassen / Einteilung	Festigkeitsklassen und Rohdichte	Gasschweißen/Autogenschweißen
Expositionsklassen / Kombinationen	Festigkeitsklassen von Bauholz für tragende Zwecke	Gebrauchstauglichkeit
Expositionsklassen / XA	Festigkeitsverhalten	Gefährdungsklasse 0
Expositionsklassen / XC	Feuchteabhängigkeit der Festigkeiten	Gefährdungsklasse 1
Expositionsklassen / XD	Feuerschutz	Gefährdungsklasse 2
Expositionsklassen / XF	Feuerschutzmittel	Gefährdungsklasse 3
Expositionsklassen / XM	Feuerwiderstand nach DIN 4102	Gefährdungsklasse 4
Expositionsklassen / XS	Feuerwiderstandsklassen	Gefährdungsklassen von Holzbauteilen nach DIN 68 800
Expositionsklassen hinsichtlich Betonangriff	Fichte – Eigenschaften – Beständigkeit – Verwendung	Gefährliche Inhibitoren
Expositionsklassen hinsichtlich Bewehrungskorrosion	Flacherzeugnisse	Gefälleestrich
FVW – Werkstoffe für Faser und Matrix	Flächenkorrosion/atmosphärische Korrosion	Gefahr für Verbindungsmittel
Falzen	Flammlöten	Gefahren von Klebstoffen für den Menschen
Fangkorb für Kohlendioxid	Flash in PP	Gefügearten des Zementits
Farbeindringverfahren	Fließbereich	Generalkonferenz für Maß und Gewichte
Faserarten	Fliesspressen	Gesamtverformung von Kunststoffen
Faserbeton / Anisotropie	Flüssige Aufkohlungsmittel	Geschichte der Metalle von der Steinzeit zur Neuzeit (1)
Faserbeton / Anwendungsgebiete	Flüssigkeitstransport in porösen Stoffen	Geschichte der Metalle von der Steinzeit zur Neuzeit (2)
Faserbeton / Begriffe	Fluoropolymerfolien	Geschichte der Metalle von der Steinzeit zur Neuzeit (3)
Faserbeton / Einleitung	Flussmittel	
Faserbeton / Faseranordnung		
Faserbeton / Fasergehalt		
Faserbeton / Faserlänge und Faserdurchmesser		
Faserbeton / Fasern		
Faserbeton / Fasern natürlicher Herkunft		
Faserbeton / Glasfaser-Rovings		

Geschichte der Metalle von der Steinzeit zur Neuzeit (4)	Glasherstellung von Antike bis zum 17. Jhd.	Haftklebstoffe
Geschichte der Produktion von Eisen, Gusseisen und Stahl	Glasscheiben-Einklebung – Direct Glazing	Halbautomatisches Schweißen
Geschichte der Textilien und Membranen	Gleichgewicht	Halbzeug und Fertigerzeugnisse
Geschichte des Betonbaus (1)	Gleichgewichtsfeuchte von Holz	Hall-Petch-Gleichung
Geschichte des Betonbaus (2)	Gleichgewichtszustandsdiagramme	Haltepunkt
Geschichte des Betonbaus (3)	Gleichmaßdehnung (Ag) beim kaltverformten Stahl	Hamburger Fernmeldeturm (272 m)
Geschichte des Betonbaus (4)	Gleichmaßdehnung (Ag) beim warmverformten Stahl	Haptisches Empfinden – Kinderspielzeug aus Holz
Geschichte des Betonbaus (5)	Gleichmaßdehnung (Ag)	Hartbitumen
Geschichte des Betonbaus – Zeitschiene	Gleichtgewichtsversagen	Hartlötten
Geschichte des Zements	Gleitbehinderungen	Hartstoffestrich
Geschichtliche Entwicklung von Betonstahl	Gleitebenen	Hauptarten des Zements
Geschlossene Zellen – Konstruktion	Gleitebenen und Gleitrichtungen	Hauptbestandteile
Geschlossene Zellen – Kunststoffe	Gleitrichtungen	Hauptbestandteile des Zement
Geschweißte Verbindungsmittel aus Stahl	Gleitsysteme im kubisch flächenz. System	Hauptquantenzahl n
Geschwindigkeit – eine vektorielle Größe	Gleitsysteme im kubisch raumz. System	hdp-Struktur
Geschwindigkeits- und Temperaturabhängigkeit	Glühen	Heizestrich
Gesetz der abgewandten Hebelarme (Waage-Regel)	Glühfarben	Herleitung des Hebelgesetzes für den Gehalt an fester Phase
Gesetzliche Regelungen zum Brandschutz	Grenzbeanspruchungen für linearelastisches Werkstoffgesetz	Herleitung des Hebelgesetzes für den Gehalt an flüssiger Phase
Gesetzlicher Rahmen	Grenzhöhe H_u	Hermesturm
Gestaltungsmittel Fuge	Grenzspannweite L_u	Herstellen von Asphaltmischungen
Gewaltbruch	Grenzspannweite eines Seils $L_{s,u}$	Herstellen von Schäumen
Gewebe	Grobkornglühen (= Hochglühen)	Herstellung der Probekörper – Proportionalstäbe
Gewebearten	Größenordnungen der verschiedenen Effekte	Herstellung des Betons
Gewinnung	Großwinkelnkorngrenzen	Herstellung und Anwendung von Gusseisen mit Kugelgraphit
Gewinnung der Rohstoffe	Grund- und Trocknungskriechen	Herstellung und Anwendung von Gusseisen mit Lamellengraphit
Gewinnung von Naturstein	Grundbruch	Herstellung und Anwendung von Weißem und Schwarzem Temperguss
Gewinnung von Rohstoffen	Grundlagen Glas – Eine Einführung	Herstellung von Eisen und Stahl im 18. Jahrhundert
Gibbssche Phasenregel	Grundlagen von Textilien und Membranen	Herstellung von Hüttensteinen
Gichtgas	Grundlagen zum Aufbau von Kunststoffen	Herstellung von Kalksandsteinen
Gitterebene	Grundlagen zum Herstellen von geformten Kunststoffteilen	Herstellung von Mauersteinen aus Normal- und Leichtbeton
Gittergerade	Grundlegende bzw. häufig auftretende Beanspruchungsfälle	Herstellung von Porenbetonsteinen
Gittergeraden und Gitterebenen	Grundprinzipien des Konstruktiven Holzschutzes – Begriffsklärung	Herstellverfahren für Betonstahl
Gitterkonstanten	Grundstähle	Hexagonal dichteste Packung (hdp)
Gitterträger	Gütegruppen	Hilfsstoffe für die Verarbeitung
Glas	Gussasphaltestrich	Hinweise zur Tragfähigkeit
Glas – Im Barock	Gusseisen	Historische Bestimmung der Lichtgeschwindigkeit
Glas in der Architektur	Gusseisen mit Kugelgraphit	Historische Holzbrücken (1)
Glas- Die totale Transparenz	Gusseisen mit Lamellengraphit (GJL)	Historische Holzbrücken (2)
Glas- Klimahüllen	Gusseisen mit Vermiculargraphit	Hochhäuser
Glas- Neue Bauaufgaben	Gussform-Verbindungselemente	Hochlegierte Stähle
Glas- Vorhangfassaden	Gusslegierungen	Hochleistungsbeton im Brückenbau
Glasanwendung im Bauwesen	Gusswerkstoffe – Eisen-Kohlenstoff-Diagramm	Hochofenprozess
Glasarten und Glasprodukte	Hängebrücken	Hochtemperaturlötten
Glaseigenschaften	Härten	Hohlraumfrei
Glasherstellung		Holzbearbeitung seit der Industrialisierung – Biegeversuche
Glasherstellung vom 18. bis zum 21. Jhd.		Holzfeuchte
		Holzschutzmittel-Druckverfahren

Holzschutzmittel-Einbringverfahren	Ionisierungsenergie	Konsistenz und Fließmittel
Holzschutzmittel-Nichtdruckverfahren	Ionisierungsenergie und Elektronenaffinität	Konsistenz und Temperatur
Holzschutzmittel-Produktgruppen	Jahreszeitliche Temperaturschwankungen	Konsistenzbereiche
Holzskelettbau	Jeffrey's Körper	Konsistenzbestimmung
Holzskelettbau – Auflager von Deckenbalken	John Hancock Center	Konsistenzkorrektur
Holzskelettbau – Diagonalaussteifungen des Tragsystems	Kalthärtende Einkomponentenklebstoffe	Konstruktion Hängebrücke
Holzskelettbau – Fußpunkt, Wandaufbau, Traufe	Kaltverformung	Konstruktionen
Holzskelettbau – Wandschnitte	Kaltwalzen	Konstruktionsarten von Membranbauwerken
Holztafelbau	Kaltziehen	Konstruktionsdetails im Holzbau – Pfosten-Riegel-Anschlüsse
Holzverbindungen	Kantenbearbeitung von Glas	Konstruktionsgrenzen
Holzverbundkonstruktionen	Kapillare Steighöhe	Konstruktionsprinzipien von Textilien und Membranen
Holzverbundquerschnitt – Brettschichtholz (BSH)	Katalysatorgifte und Eindringen des Wasserstoffs	Konstruktive Maßnahmen
Holzverbundquerschnitt – Spanplatte	Kathodische Fremdmallüberzüge	Konstruktiver Holzschutz – Gefährungsklassen (GK)
Holzverbundquerschnitt – Sperrholzplatte	Kathodische Teilreaktion	Konstruktiver Holzschutz – Schutz gegen Feuchteleitung
Holzverfärbende Pilze	Kathodischer Korrosionsschutz: Methoden	Konstruktiver Holzschutz – Schutz gegen Niederschläge
Holzwerkstoffklassen	Kathodischer Korrosionsschutz: Prinzip	Konstruktiver Holzschutz – Tauwasser Austritt diffusionsoffener Bauteile (2)
Holzerstörende Insekten	Kelvin-Körper	Konstruktiver Holzschutz – Tauwasser-schutz (1)
Holzerstörende Pilze	Kenngößen der Viskosität	Konstruktiver Holzschutz – feuchtege-fährdete Holzbauteile
Home Insurance Building	Kenngößen feuerschutztechnische Berechnung – ummantelter Stahl	Kontaktkleber
Homogene Zustandsfelder	Kennzeichnung der Betonstahlerzeugnisse	Kontaktkorrosion (Bimetallkorrosion)
Hooke'sches Grundelement	Kerbschlagarbeit	Koordinationszahl KZ
Hookesches Gesetz	Kerbschlagbiegeversuch	Korngrenzen
Hundsche Regel	Kerbschlagbiegeversuch (DIN EN 10 045)	Korngrößenverteilung des minerali-schen Zuschlags – Fullerkurve
Hybride Bauteile	Kerbschlagzähigkeit	Korrosion der Bewehrung
Hybrider Brückenträger	Kernholz und Splintholz	Korrosion der Oberfläche (Betonkorro-sion) des Bauteils
Hybrider Rohrquerschnitt	kfz-Struktur	Korrosion: Allgemeine Definition
Hydratation	Kiefer – Eigenschaften – Beständigkeit – Verwendung	Korrosionsarten
Hydratationsbeeinflussende Stoffe	Kiesnester	Korrosionselement
Hydratationsgrad	Kinematik	Kosten und Finanzierung
Hydratationsstufe I	Kippen	Kriechen von Beton
Hydratationsstufe II	Klassifizieren von Bitumen	Kristallite
Hydratationsstufe III	Kleben	Kristallsysteme
Hydratationsstufen	Kleblächenvorbehandlung	Kriterium für Art der Bindung
Hydratationsverlauf	Kleblösungen	krz-Struktur
Hydratationswärme	Klebstoffe und der Vorgang des Klebens	Kubisch flächenzentriert (kfz)
Hydratphasen	Kleinwinkelkorngrenzen	Kubisch primitiv (kp)
Hypothese von Dalton	Klinkerminerale	Kubisch raumzentriert (krz)
Im Bauwesen verwendete Holzproduk-te	Knetlegierungen	Künstlich eingeführte Luftporen
Indirekte Lasteinwirkungen	Knicken	Kunststoffe
Induktionshärten	Knickpunkt	Kunststoffe – Einführung
Induktionslötten	Kohärente und Inkohärente Größen	Kunststoffproduktion
Infrarot	Kohlenstoffäquivalent	Kunststoffüberzüge
Inhibitoren	Kolbenlötten	Kupfer
Inhomogene Zustandsfelder	Kombinationen von Werkstoffen	Kupfer-Legierungen
Innenrüttler	Komponenten der van der Waals-Kräfte	Kupferprodukte im Bauwesen
Integration von Formteilen	Komponentenbestimmung	Kuppelkonstruktionen
Interkristalline Spannungsrissskorrosion	Konsistenz (des Frischbetons)	
Intermetallische Verbindungen		
Ionenbindung am Beispiel des Kalium-chlorid		
Ionenradien		

Kurze geschichtliche Entwicklung	Magmatite	Membrankonstruktionen
Kurzer geschichtlicher Abriss	Magnesiaestrich	Membranränder
Kurzzeichen für Kunststoffe	Magnetische Quantenzahl	Messung der Lichtgeschwindigkeit auf der Erde
Kurzzeitfestigkeiten von Vollholz	Magnetpulververfahren	Messung des pH-Wertes
Lärche – Allgemeine Eigenschaften	Mahlen des Zement	Metall-Aktivgas/Inertgas-Schweißen (MAG/MIG)
Lagermatten	Mahlen des Zements	Metall-Lichtbogen-Schweißen
Lagern, Abfüllen, Transport	Mahlverfahren	Metall-Schutzgas-Schweißen (MSG)
Lagerung und Lieferung von Betonstahl	Makroskopischer Aufbau von Holz	Metall im Verbundbau
Langerzeugnisse	Makroskopisches Bruchverhalten	Metallische Abdichtungen
Langzeitverformungen von Holz – Allgemeines	Mannheimer Fernmeldeturm (204 m)	Metallische Fassadenbekleidung
Lastabtrag bei Balkenbrücken	Manuelles Schweißen	Metallische Oberflächen (1)
Lastabtrag der Bogenbrücken	Marina City	Metallische Oberflächen (2)
Lastansätze für textile Bauwerke	Martensit	Metallische Oberflächen (3)
Lastunabhängige Verformung (Schwinden und Quellen)	Maschinelle Sortierklassen von Schnittholz	Metallische Oberflächen (4)
Laubholz	Massebilanz	Metamorphite
Ledeburit	Maßnahmen des Holzschutzes	Metastabiles System (Fe-Fe ₃ C-Diagramm)
Leerstelle	Materialausdruck und Wahrnehmung von Mauerwerk	Mikroskopischer Aufbau von Holz
Legieren/Legierungselemente	Materialbruch	Mikroskopisches Bruchverhalten
Legierte Edelstähle	Materialempfinden und Formensprache namhafter Baumeister (1)	Milikanversuch
Legierte Qualitätsstähle	Materialempfinden und Formensprache namhafter Baumeister (2)	Millersche Indizes für Hexagonale Kristalle
Legierungen	Materialien	Minimalfläche
Leichtbeton / Definitionen	Mathematisch-numerische Methode	Mischanlage
Leichtbetonarten	Matrixwerkstoffe	Mischungsentwurf
Leichtestrich	Mauerwerk	Mischungsstabilität
Leimbauweisen	Mauerwerk in der Architektur	Modellkörper
Leime, Leimlösungen	Mauerwerk- Rom und Byzanz	Moderne Holzverb. mit Stahlelementen
Lesbarkeit der Lastabtragung	Mauerwerk- Technische Entwicklung in der griechischen Antike	Modifizierte Betrachtung nach IRWIN und Bruchzähigkeit
Lethersich-Körper	Mauerwerk- Technische Entwicklung in der römischen Antike	Momentan- und Durchschnittsgeschwindigkeit
Lewis-Diagramm	Mauerwerk- Technische Entwicklung vom Mittelalter bis ins 18. Jhd.	Momenten-Normalkraft-Interaktion
Linearelastische und nichtlineare Grenzzustände	Mauerwerk- Von der griechischen zur römischen Baukunst	Monolithe – die Kraft des Einfachen
Liquiduslinie	Mauerwerk- vom Mittelalter bis zur Renaissance	Montage von Membrankonstruktionen
Listenmatten	Mauerwerksarten	Motivation
Lochbleche und Stahlblechformteile	Mauerwerksbau- Frühe Herstellungsmethoden	Nachbehandlung von Beton
Lochkorrosion	Mauerwerksbau- Von der Renaissance bis ins 18. Jahrhundert	Nachbehandlungsdauer
Löslichkeit	Mauerwerksverbände	Nachbehandlungsverfahren
Lösung für den vollplastischen Zustand	Maxwell-Körper	Nachentschwefelung
Lösung für elastisch-idealplastisches Werkstoffgesetz	Maxwellsche Wellentheorie	Nachträgliche Zugabe von Fließmitteln
Lösung für linearelastisches Werkstoffgesetz	Mechanische Beanspruchbarkeit von Naturstein	Nadelhölzer
Lösungsgleichungen	Mechanische Eigenschaften von Glas	Nagel- und Schraubverbindungen
Lösungsmittel	Mechanische Kenngrößen	Nagelblechbauweise – Greimbau
Lösungswärme	Mechanische Modelle	Nagelplatten und Nagelbleche
Lösungsweg	Mechanische Oberflächenbehandlung	Nahtformen
Löten	Mechanisches Verhalten von Werkstoffen	Naturstein im Hochbau
Löten – Lötfehler	Mehlkorngehalt	Naturstein im Straßen- und Verkehrswegebau
Löten – Rollenlöten	Mehrschige Beanspruchung	Naturstein in der Betonherstellung
Löten – Verbindungslöten		Nebenbestandteile des Zement
Lotbadtauchlöten		Nebengruppen
Lote		Nebenquantenzahl
Luftdichtheit		Nernst-Gleichung (1/2)
Luftgehalt		Nernst-Gleichung (2/2)
Luftvergütung		

Neue Möglichkeiten zur Definition des Kilogramms	Perlitbildung	Qualitätsstähle in anderen Bereichen – Einsatzstähle
Neue Wege im Schalenbau – Heinz Isler	Petronas Towers	Qualitätsstähle in anderen Bereichen – Federstahl
Neutralisation	Pfosten-Riegel-Anschlüsse – Einsatzorte	Quantenzahlen
Newton'scher Dämpfer	Pfosten-Riegel-Anschlüsse – Konstruktionsbeispiele	Quantisierung
Nichthärtende Klebstoffe	Phänomene beim kapillaren Flüssigkeitstransports	Quell- und Schwindmaße von Holz
Nichtrostende Stähle	Phosphatieren	Querkontraktionszahl
Nickel	Physikalisch abbindende Klebstoffe	Raumsegmente – Konstruktion
Niedriglegierte Stähle	Physikalisch härtende Klebstoffe	Raumsegmente – Kunststoffe
Nieten	Physikalischer Angriff auf Beton	Reaktion vom Monomeren zum Polymeren
Nieten – Allgemeines	Pigmente der Deckbeschichtung	Reaktionsklebstoffe – Zweikomponentenkleber
Nietformen: Niete, Vollniete, Hohniete	Pigmente der Grundbeschichtung	Reaktionsklebstoffe, Einkomponentenkleber
Nitrierhärten	Plastische Formänderungsenergie	Reale Kristalle
Normalglühen:	Plastisole	Reale Werkstoffgesetze
Normalkraft	Plastisolklebstoffe	Rechenwerte des E-Moduls für Beton
Normalspannungsreihe	Plastizität	Reckalterung
Nutzung der Plastizität	Plattieren	Recken
Nutzungsdauer	Plötzliches Versagen eines Stahlbetonbauteils mit zu geringer Mindestbewehrung	Reduktion CO ₂
Obere Anwendungsgrenztemperatur	Polare und unpolare Bindungen	Reflexion des Lichts
Obere Streckgrenze (ReH)	Polyaddition	Regeln zur Bestimmung der Bindigkeit
Oberfläche und Textur – Betonrezeptur	Polykondensation	Regelungstechnik
Oberfläche und Textur – Schalung (1)	Polykondensation am Bsp. von Phenolharz	Regionaltypische Entwicklungen im Fachwerkbau
Oberfläche und Textur – Schalung (2)	Polymerisation – ionisch	Reibschweißen
Oberfläche und Textur – Schalung (3)	Polymerisation – radikalisch	Reifeformel nach Saul-Nurse
Oberfläche und Textur – Sichtbeton	Porenlösung des Zementsteins	Reißlänge Ru
Oberfläche und Textur – Übersicht	Portlandzementklinker (K)	Rekristallisationsglühen
Oberflächengestaltung	Potential und Normalspannungsreihe	Relaxation und Kriechen von Kunststoffen
Oberfläche und Textur – Verfahren der Oberflächenbearbeitung	Potentialunterschied	Relevante Eigenschaften der für Bauzwecke eingesetzten Holzarten
Oberfläche und Textur von Mauerwerkskonstruktionen	Pourbaix-Diagramme für Aluminium, Zink und Chrom	Reliance Building
Oberflächen & Textur – Beschichtungen (1)	Pourbaix-Diagramme für Zinn, Kupfer und Nickel	Rezyklierbarkeit
Oberflächen & Textur – Beschichtungen (2)	Pressdübel – Einlassdübel & Einpressdübel	Rheologie
Oberflächenbehandlung	Pressdübel – Einseitiger Einpressdübel	Richtungsabhängige Festigkeiten von Vollholz
Oberflächenbehandlung von Glas	Pressdübel – Multi-Krallen-Dübel	Ring und Kugel
Oberflächenveränderung des Mauerwerks	Pressschweißen	Rissbeanspruchungsarten
Ökologie	Primärenergiebedarf	Robustheit der Konstruktion
Ölvergütung	Primärkristall	Roheisen
Oktettregel	Profilieren	Rohrleitungen und Installationszubehör
Olivetti-Ausbildungszentrum	Project Eden – Folienkissen aus Kunststoff	Rohstoffe
Optische Eigenschaften von Glas	Prüfung des Frost-Tau-Widerstandes von Beton	Rohstoffmischung
Orbitalbelegung von Argon	Pumpbeton – interaktives Bild	Rüttelbohlen – Oberflächenrüttler
Orbitalbelegung von Eisen	Punktschweißen (RP)	Rüttler
Oxidation und Reduktion	Puzzolane	SI-Basiseinheiten
Oxidationsbitumen	Qualitätsstähle	SI-Vorsätze
pH-Wert Skala	Qualitätsstähle – Anwendungsbeispiel	Säuren und Basen
pOH-Wert	Qualitätsstähle in anderen Bereichen	Sanierungsmaßnahmen
PVC- Folien	Qualitätsstähle in anderen Bereichen – Automatenstähle	Sanierungsmaßnahmen Karlshof
Passivität		Sauerstoffkorrosion
Patentieren		Schacht- und Drehrohröfen
Pauli-Prinzip		Schäden infolge mangelnder Verarbeitbarkeit
Peritektikum		
Peritektische Umsetzung		
Perlit		

Schalentragwerke	Schweißfehler – Wiedererwärmungsrisse	Spannungs-Dehnungs-Diagramm eines kaltverformten Stahls
Schalung	Schweißfehler – Wolframeinschlüsse	Spannungs-Dehnungs-Diagramm eines warmverformten Stahls
Schalung (Anforderungen)	Schweißfehler – Wurzelrisse	Spannungs-Dehnungs-Diagramme
Schlacke	Schweißnahtkorrosion	Spannungs-Dehnungs-Linie von Beton
Schlagartiges Versagen eines Stahlbetonbauteils in der Druckzone	Schweißnahtprüfung – Allgemeines	Spannungs-Dehnungs-Verhalten von Stahl mit ausgeprägter Steckgrenze
Schliffbildherstellung (geätzt)	Schwerbeton / Dichte der Zuschläge	Spannungs-Dehnungs-Verhalten von Stahl ohne ausgeprägter Steckgrenze
Schliffbildherstellung (ungeätzt)	Schwerbeton / Einführung	Spannungsarmglühen
Schluss	Schwerbeton / Herstellung	Spannungsbegrenzung
Schmelzkleber (Feste Klebstoffe)	Schwerbeton / Zuschlagarten	Spannungsintensitätsfaktor (SIF)
Schmelzklebstoffe	Schwimmender Estrich	Spannungskonzentration durch Kerben
Schmelzschiessen	Schwinddehnung / Schrumpfdehnung	Spannungsrissskorrosion
Schmelztauchüberzüge (Feuerverzinken)	Schwinden und Quellen von Holz	Sparsame Konstruktionen
Schmidtsches Schubspannungsgesetz	Schwindverformungen	Sprödes Bauteilversagen
Schmiedeeisen- Das Prinzip der Unter- spannung	Seagram Building	Sprödes Werkstoffversagen
Schmieden	Sears Tower	St.-Venant-Element
Schneelasten	Sedimente	Stabdübel- und Bolzenverbindungen
Schneiden von Glas	Sekundärmetallurgie	Stabiles System (Fe-C-Diagramm)
Schnellarbeitsstähle	Selbsttragende Pneus – Konstruktion	Stabilitätsversagen
Schnelle Hydratationswärmeentwicklung	Selbsttragende Pneus – Kunststoffe	Stähle für den Stahlbau
Schnittholzquerschnitt – Balken	Sichtbeton / Anforderungen	Stähle mit einem Kohlenstoffanteil von mehr als 0,6 M.-% – Ein kurzer Einblick
Schnittholzquerschnitt – Bohle	Sichtbeton / Einführung	Stähle mit einem Kohlenstoffgehalt weniger 0,2 M.-% – Ein kurzer Einblick
Schnittholzquerschnitt – Brett	Sieblinien	Ständige Lasten
Schnittholzquerschnitt – Kantholz	Siemens-Martin-Verfahren	Stahlbaustähle – Allgemeines
Schnittholzquerschnitt – Latte	Sinnlichkeit – Abschirmung elektromagnetischer Strahlen	Stahlbaustähle – Feinkornbaustähle
Schrägseilbrücken	Sinnlichkeit – Geräuschempfinden	Stahlbaustähle – Vergütungsstähle
Schrauben	Sinnlichkeit – Geruch	Stahlbegleiter
Schrauben und Zubehör	Sinnlichkeit – Wärme (1)	Stahlerzeugung heute
Schraubenversetzung	Sinnlichkeit – Wärme (2)	Stahlguss
Schrödinger Gleichung	Sinnlichkeit – haptisches Empfinden	Stahlherstellung
Schub- und Scherbeanspruchung	Sinnlichkeit des Mauerwerks	Stahlschlüssel
Schubspannungen	Sinnvolle Kombination von Werkstoffen – Betonkonstruktionen	Stahlverbundkonstruktionen – Entwicklung des Verbundbaus
Schutzgase	Sinnvolle Kombination von Werkstoffen – Holzkonstruktionen	Stahlzugversuch
Schwarzer Temperguss	Sinnvolle Kombination von Werkstoffen – Stahlkonstruktionen	Statik von Membrankonstruktionen
Schwefeldioxid	Skylontower (160 m)	Staubpartikel
Schweißbeugung	Slumpmaßklassen	Steifigkeitseigenschaften der Festigkeitsklassen von Bauholz für tragende Zwecke
Schweißbeugung – Kohlenstoffäquivalent CEv	Solare Heizunterstützung	Steinarten und –bezeichnungen (Hüttensteine)
Schweißelektroden – Allgemeines	Soliduslinie	Steinarten und –bezeichnungen (Kalksandsteine)
Schweißelektroden – Elektrodentypen	Sondergusseisen	Steinarten und –bezeichnungen (Mauersteine aus Normal- und Leichtbeton)
Schweißen	Sonneneinstrahlung	Steinarten und –bezeichnungen (Porenbetonsteine)
Schweißen – Allgemeines	Sonstige Anwendungen von Metallen im Bauwesen	Steinformate
Schweißen hochlegierter Stähle	Sonstige Verwendung	Steinoberflächen
Schweißen von Betonstahl	Sorptionsisotherme und Mechanismen der Feuchteaufnahme	Stellung der Metalle im Periodensystem
Schweißfehler – Einbrandkerben	Space Needle in Seattle	Stoffkreislauf
Schweißfehler – Endkraterisse	Spanndrahtlitzen	
Schweißfehler – Fehler im Schweißnahtübergang	Spannstähle	
Schweißfehler – Gaseinschlüsse oder Poren	Spannstahlbezeichnung	
Schweißfehler – Kaltrisse	Spannstahlverankerungen	
Schweißfehler – Lamellenrisse	Spannungen	
Schweißfehler – Schlackeneinschlüsse	Spannungs-Dehnungs- und Festigkeitsverhalten von Werkstoffen	
Schweißfehler – Warmrisse		

Stoffraumrechnung	Türme	Valenzschale
Strahlenhärtbare Klebstoffe	Türme für Windkraftanlagen	van der Waals-Kräfte
Strangguss	Two Union Square	Variation der Netzkettlänge bei vernetzten Kunststoffen
Stromquellen	Typische Werte der kritischen Schubspannung von Einkristallen	Variation des Molekulargewichtes
Struktur von Aluminium	Überblick Fügetechniken	Veränderliche Lasten
Struktur von Kupfer	Überblick über die im Bauwesen verwendeten Holzarten	Veränderung der chemischen Eigenschaften
Struktur von Zink	Überblick über die im Bauwesen verwendeten Holzprodukte	Verarbeitbarkeit: Allgemeines
Stufenversetzung	Überseutektoider Stahl	Verarbeiten der "polyadditiven" Duroplaste und der Elastomere
Stuttgarter Fernsehturm	Überseeische Holzarten	Verarbeiten der Duroplaste-Polykondensationsprodukte
Substitutionsatom	Übersicht NE-Metalle	Verarbeiten der Thermoplaste
Synklastische Formen	Übersicht Zementproduktion	Verbindungsmittel – Moderne Verbindungstechnik mit Nagelplatten
Systematik der Klebstoffe in Anlehnung an DIN 16 920	Übersicht des Einflusses der Legierungselemente auf die Schweißbarkeit	Verbindungsmittel – Moderne Verbindungstechnik mit Stahlelementen
Tanne – Eigenschaften – Beständigkeit – Verwendung	Übersicht häufiger Kunststoff-Anwendungen im Bauwesen	Verbindungsschweißen
Tauchhärten	Übersicht möglicher Kunststoff-Anwendungen im Bauwesen	Verbund im Biegebalken
Technische Elastizitätsgrenze (Rp0,01)	Übersicht von verschiedenen Stoffen und deren Dichten	Verbundestrich
Temperaturabhängiges mechanisches Verhalten der Kunststoffe	Übersicht wichtiger Nichteisenmetalle (1)	Verbundwerkstoffe im Bauwesen
Temperaturabhängigkeit der Holzfestigkeit	Übersicht wichtiger Nichteisenmetalle (2)	Verdichtungsmaßklassen
Temperaturabhängigkeit der Kunststoffef	Ultrahochfester Beton im Brückenbau	Verdichtungsversuch
Temperaturbedingte Längenänderungen von Holzbauteilen	Ultraschallverfahren	Verdrehen (Torsion)
Tempem	Ultraviolett	Verdrillen- Tordieren
Thermische Behandlung von Glas	Unlegierte Edelstähle	Verfahren der Formgebung Faserverstärkter Kunststoffe
Thermische Eigenschaften von Glas	Unlegierte Qualitätsstähle	Verfahren der Zementherstellung
Thermisches Verhalten von Naturstein	Unlegierte Stähle	Verfahren zur Herstellung von Nietverbindungen
Thermoelaste	Unlegierte Stähle mit Kohlenstoffanteil von 0,2 bis 0,6 M.-% – Eine kurzer Einblick	Verfahrensbedingte Fehler
Thermoplaste, Duroplaste und Elastomere	Unlegierte Stähle mit Mn < 1%	Verfestigungsbereich
Thermoplastische Elastomere (in DIN 7724 nicht erfasst)	Unlegierte Stähle mit Mn > 1% und legierte Stähle mit Legierungsgehalten < 5% (niedrig legierte Stähle)	Verfestigungsbereich beim kaltverformten Stahl
Thermoplastische Kunststoffe (Thermoplaste)	Unschärferelation	Verfestigungsbereich beim warmverformten Stahl
Thomson-Kathodenstrahl Experiment	Untere Streckgrenze (ReL)	Verfestigungsmechanismen
Tonprobe-Holztreppe	Untereutektoider Stahl	Verformung infolge Trocknungsschwinden für Normalbeton
Tour d' Orientation	Unterpulverschweißen (UP)	Verformungen
Trägereinlagen von Bitumenbahnen	Unterschiede bei zugänglichen Poren	Verformungen des Risses
Tragfähigkeit von Konstruktionen	Unterschiedliche Diffusionsmechanismen	Verformungen durch Kriechen
Tragprinzip von Membrankonstruktionen	Untersuchen und Klassifizieren von Bitumen	Verformungen durch Schwinden
Tragwerke im Hochbau	Untersuchungsverfahren	Verformungen durch Setzungen
Tragwerke im Wasserbau und Grundbau	Unterteilung der Bindungen	Verformungen durch Temperatureinwirkungen
Transamerica Pyramid	Unterteilung in Haupt- und Nebengruppen	Verformungsarbeit
Transkristalline Spannungsrisskorrosion	Unterteilung in Metalle und Nichtmetalle	Verformungskennwerte
Transmission des Lichts	Urformen von Thermoplasten (Plastomere)	Verformungsverhalten in Abhängigkeit der Richtung des Lastangriffes
Transport des Betons	Urformen von duroplastischen Kunststoffen	Verformungsverhalten von Holz
Transportbeton	Vakuumbehandlung	Verformungsverhalten von Holz unter Kurzzeitbelastung
Trianon		Vergleich Bezeichnungssystem mit Kurzname und Werkstoffnummer
Tripelpunkt des Wassers und absoluter Nullpunkt		Vergleich alte und neue Norm
Trockenestrich		Vergleich zwischen Normalbeton und Hochfestem Beton
Trocknen von Holz		
Trocknungsschwinddehnung		

Vergleich: Polymerisation-Polykondensation-Polyaddition	Von den Anfängen bis zum mod. Holzbau (4)	Widerstandsklassen
Vergleichende Gegenüberstellung der Querschnittsabmessungen von Schnitthölzern	Von der Ionenbindung zur Atombindung	Widerstandslöten
Vergüten	Voraussetzungen für die Bildung eines Korrosionselements	Winderhitzer (Cowper)
Vergütungsschaubild	Vorbehandlung der Metalloberflächen	Windlasten
Vergütungsverfahren	Vorspannung von Textilien und Membranen	Winkel
Verkehrslasten	Vulkanisation	Wirksames Betonalter
Verlegen von Bitumenbahnen	Vébemaßklassen	Wirkung äußerer Kräfte auf Tragwerke
Verpressdübel (1)	WIBA-NET	Wirkung des eingedrungenen Wasserstoffs
Verpressdübel (2)	Wärmedämmung	Wirkungsgruppen und Kennzeichnung der Betonzusatzmittel
Versagensarten	Wärmeeinflusszone (WEZ)	Wirtschaftlichkeit
Verschiebung der Neutralen Faser bei nichtlinearem Werkstoffgesetz	Walzen	Witterungsbeständigkeit von Naturstein
Verschiedene Betonstahlmatten	Walzgerüste	Wolfram-Inertgas-Schweißen(WIG)
Versetzungsdichte	Walzstraßen	Woolworth Building
Verspannen, Verzerren und Verbiegen der Gleitpakete	Wanddämmung	World Trade Center
Verwendete Leime	Warmwalzen	ZTA-Schaubilder)
Verwendungsgeschichte von Eisen und Stahl	Wash-Primer	ZTU-Schaubilder)
Verwendungsgeschichte von Glas	Wassereindringwiderstand	Zangenanschlüsse – Einsatzorte
Verwendungsgeschichte von Textilien und Membranen	Wasserstoff- und Sauerstoffelektrode	Zeichnungsmatten
Verzögert-Elastisches Verhalten	Wasserstoffentstehung bei der Korrosion	Zeit-Temperatur-Schaubilder
Video: Betoneinbau	Wasserstoffkorrosion	Zeitabhängige Verformungen – Berechnungsgrundlagen
Viskoelastizität	Wasserstoffversprödung (kathodische Spannungsrißkorrosion)	Zeitpunkt der Konsistenzbeurteilung
Viskosität	Wassertropfenkorrosion	Zeitraum (der Konsistenz)
Visuelle Prüfung	Wasservergütung	Zeitraum der Konsistenzbeibehaltung
Visuelle Sortierklassen von Schnittholz	Wechselwirkung fluktuierender Dipol – induzierter Dipol	Zeitschiene der Metall-Geschichte von der Steinzeit zur Neuzeit
Vollautomatisches Schweißen	Wechselwirkung permanenter Dipol – induzierter Dipol	Zeitschiene der Entwicklung des Holzbaus
Vollkommen unlöslich im flüssigen und im festen Zustand	Wechselwirkung permanenter Dipol – permanenter Dipol	Zeitschiene der geschichtlichen Entwicklung von Kunststoffe
Vollkommene Löslichkeit im flüssigen Zustand und beschränkte Löslichkeit im festen Zustand	Wechselwirkung von Schub- und Normalspannung	Zeitverformungslinie
Vollkommene Löslichkeit im flüssigen Zustand, vollkommene Unlöslichkeit im festen Zustand	Wechselwirkung zwischen Porosität und Baustoffeigenschaften	Zement
Vollkommene Löslichkeit im flüssigen und im festen Zustand	Weichglühen	Zement im Altertum – Opus Caementitium
Vom Brückenpfeiler zum Turm	Weichlöten	Zement in der Neuzeit – moderne Zemente
Vom Eisen zum Stahl- Fachwerke und erste Skelettbauten	Weißer Temperguss	Zementestrich
Vom schwarzen Bakelit zu transparenten High-Tech-Kunststoffen (1)	Weiten und Engen	Zementhaltige Überzüge
Vom schwarzen Bakelit zu transparenten High-Tech-Kunststoffen (2)	Weitere Arten von Betonsteinen	Zementit
Vom schwarzen Bakelit zu transparenten High-Tech-Kunststoffen (3)	Wellen-/ Schwalllöten	Zementsäcke
Vom schwarzen Bakelit zu transparenten High-Tech-Kunststoffen (4)	Wellenfunktion	Zener-Körper
Von den Anfängen bis zum mod. Holzbau (1)	Werkstoffbedingte Fehler	Zentrische Zugfestigkeit
Von den Anfängen bis zum mod. Holzbau (2)	Werkstoffbezeichnungen des Gusseisens mit Kugelgraphit	Ziehen
Von den Anfängen bis zum mod. Holzbau (3)	Werkstoffbezeichnungen des Gusseisens mit Lamellengraphit	Ziel des Zugversuchs
	Werkstoffe für Niete	Zink und Kupfer in fließendem Wasser
	Werkstoffe zur Verklebung mit Zweikomponentenklebern	Zink-Legierungen
	Werkstoffgesetz	Zinkprodukte im Bauwesen
	Werkstoffübergang	Zug-, Biege-, Druck- und Schubfestigkeit
	Werkstoffverhalten bei Dehnungsumkehr und Einteilung der Werkstoffgesetze	Zugabewasser
		Zugängliche und nicht zugängliche Poren
		Zugfestigkeit
		Zugversuch
		Zunder

Zur Konstruktion der Schrägseilbrücken	Zusammenstellung der technischen Kennwerte von Nadelhölzern	Zuschlag / Stoffe organischen Ursprungs
Zusammenfassung: Stahlherstellung	Zusatzmittel	Zuschnitt von Textilien und Membranen
Zusammenhang zwischen Zug- und Druckfestigkeit	Zusatzstoffe	Zuschnittsarten
Zusammensetzung und Aufbau von Holz	Zuschlag	Zustandsbereiche von Kunststoffen
Zusammensetzung und Verarbeitung des Betons (Zusatzstoffe, Verdichtung)	Zuschlag / Abschlämbbare Bestandteile	Zwangsbeanspruchungen
Zusammensetzung von Zement	Zuschlag / Alkalilösliche Kieselsäure	Zweifamilienhaus (Baujahr 1975)
Zusammenstellung der technischen Kennwerte von Laubhölzern	Zuschlag / Erhärtungsstörende Stoffe	Zweikomponentenklebstoffe
	Zuschlag / Schädliche Bestandteile	Zwischengitteratom
	Zuschlag / Schwefelverbindungen	
	Zuschlag / Stahlangreifende Stoffe (Salze)	

Anhang G: Ausgewählte Seiten des Lehrpfads „Konsistenzbestimmung“

Im folgenden sind ausgewählte Seiten des WiBA-Net-Lehrpfads „Konsistenzbestimmung“ dargestellt. Dies sind die Seiten des Clusters „Versuche zur Konsistenzbestimmung“. Die Reihenfolge der Darstellung entspricht der Reihenfolge im Lehrpfad. Zu jeder Seite ist vermerkt, ob es sich um eine Haupt- oder Nebenseite handelt, bei Nebenseiten wird die zugehörige Haupt- bzw. übergeordnete Nebenseite mit angegeben. Eine Übersicht der Struktur des Lehrpfads kann Tabelle 3.2 entnommen werden.

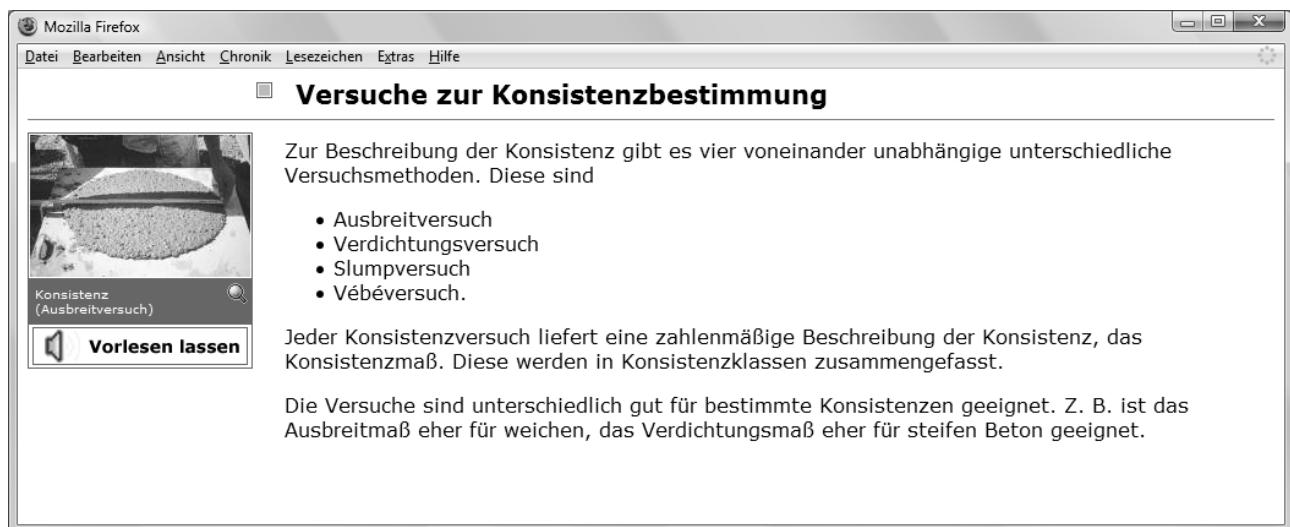


Abbildung G.1: Seite „Versuche zur Konsistenzbestimmung“ (Hauptseite)



Abbildung G.2: Seite „Bestimmung des Ausbreitmaßes“
(Nebenseite 1. Ordnung zur Hauptseite „Versuche zur Konsistenzbestimmung“)

Ausbreitmaßklassen

Vorlesen lassen

Um den Beton hinsichtlich seiner Verarbeitbarkeit zu charakterisieren, hat sich bei Ausbreitmaß und Verdichtungsmaß die Einführung von Konsistenzbereichen als zweckmäßig erwiesen.

Der Konsistenzbereich **weich** wird auch als Regelkonsistenz bezeichnet, denn dieser Beton wird wegen seiner leichten Verdichtbarkeit im allgemeinen auch bei Bauteilen eingesetzt, wo im Grunde auch eine steifere Konsistenz ausreichen würde.

Konsistenzbereiche	Konsistenzklasse	Ausbreitmaß f [mm]
sehr steif	-	-
steif	F1	< 340
plastisch	F2	350 - 410
weich	F3	420 - 480
sehr weich	F4	490 - 550
fließfähig	F5	560 - 620
sehr fließfähig	F6	> 630

Abbildung G.3: Seite „Ausbreitmaßklassen“
(Nebenseite 2. Ordnung zur Nebenseite „Bestimmung des Ausbreitmaßes“)

Vergleich der Konsistenzklassen

In der unten stehenden Tabelle werden die verschiedenen Konsistenzklassen gegenübergestellt.

Vergleich der Konsistenzklassen								
Konsistenzbereiche	Ausbreitmaß		Verdichtungsmaß		Slumpmaß		Vébémaß	
	Konsistenzklasse	Ausbreitmaß f [mm]	Konsistenzklasse	Verdichtungsmaß c [-]	Setzmaß s [mm]	Konsistenzklasse	Setzzeit t_z [s]	Konsistenzklasse
sehr steif	-	-	C0	> 1,46				
steif	F1	< 340	C1	1,45 - 1,26				
plastisch	F2	350 - 410	C2	1,25 - 1,11	10 - 40	S1	> 31	V0
weich	F3	420 - 480	C3	1,10 - 1,04	50 - 90	S2	31 - 21	V1
sehr weich	F4	490 - 550	-	-	100 - 150	S3	20 - 11	V2
fließfähig	F5	560 - 620	-	-	160 - 210	S4	10 - 6	V3
sehr fließfähig	F6	> 630	-	-	≥ 220	S5	5 - 3	V4

Abbildung G.4: Seite „Vergleich der Konsistenzklassen“
(Nebenseite 2. Ordnung zur Nebenseite „Bestimmung des Ausbreitmaßes“)

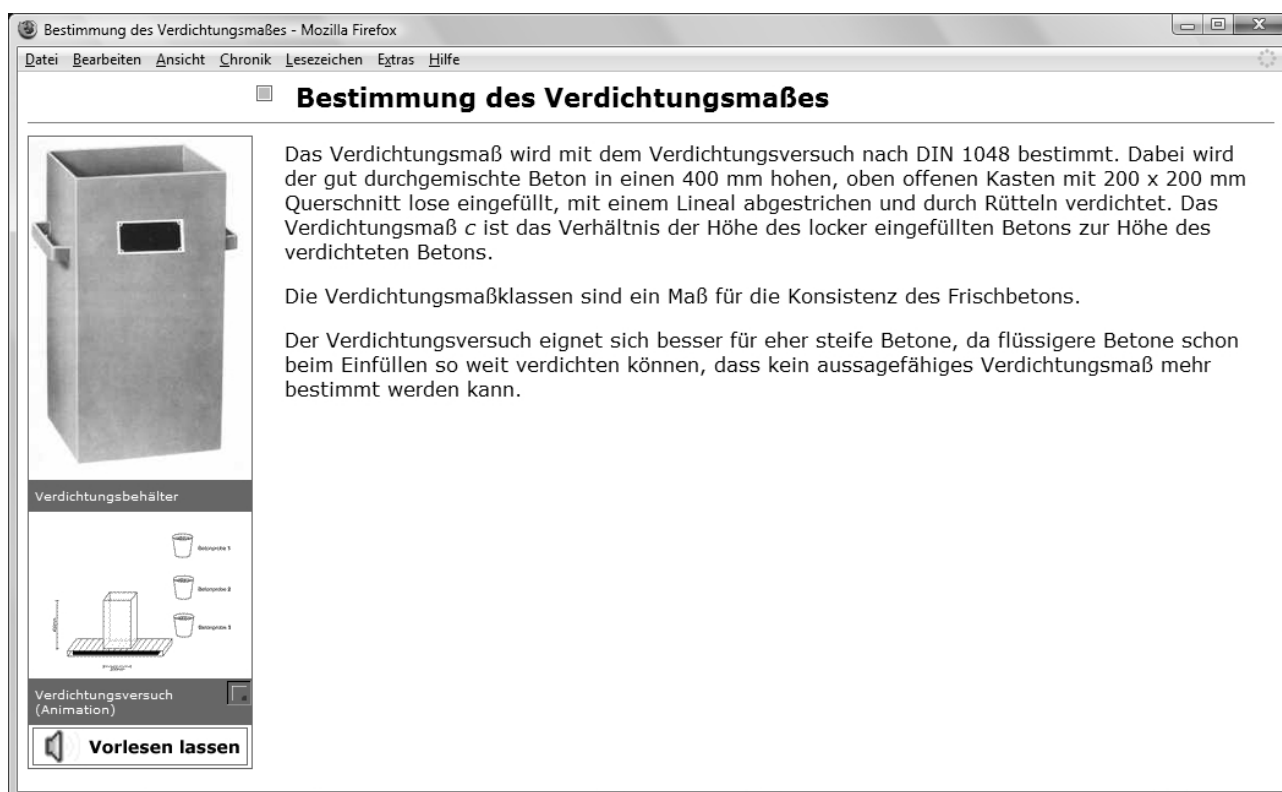


Abbildung G.5: Seite „Bestimmung des Verdichtungsmaßes“
(Nebenseite 1. Ordnung zur Hauptseite „Versuche zur Konsistenzbestimmung“)

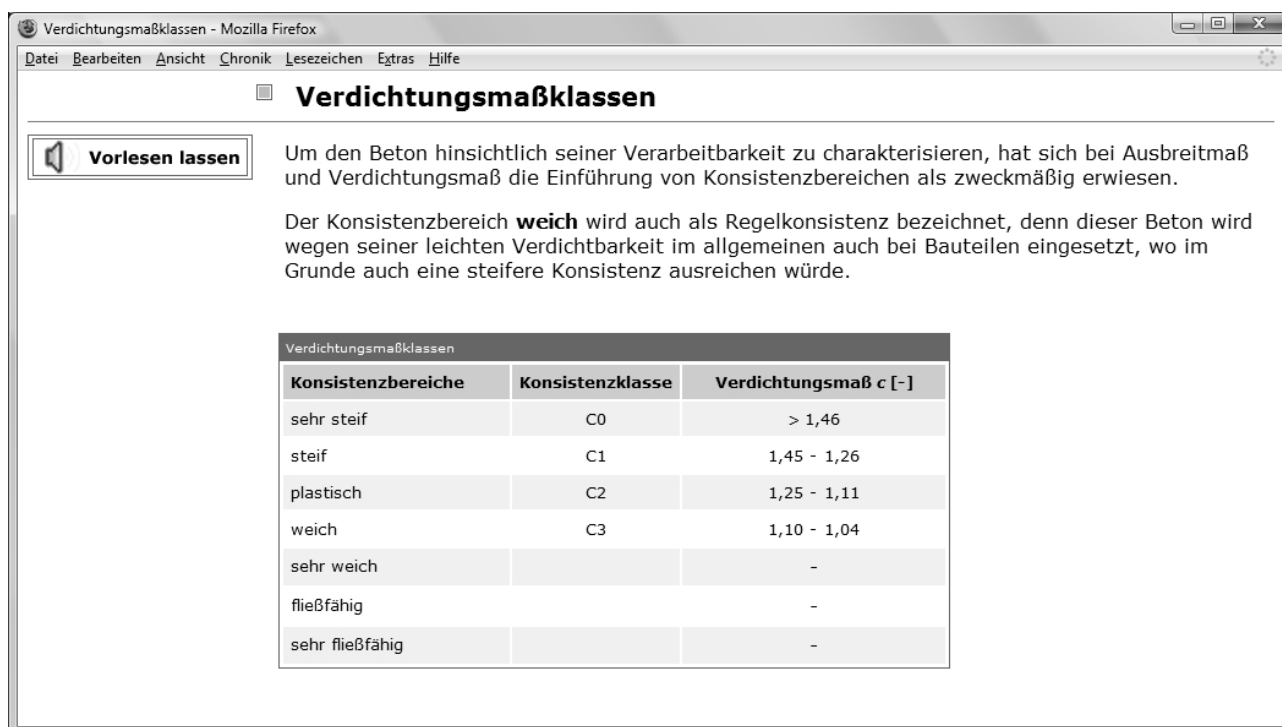


Abbildung G.6: Seite „Verdichtungsmaßklassen“
(Nebenseite 2. Ordnung zur Nebenseite „Bestimmung des Verdichtungsmaßes“)



Abbildung G.7: Seite „Bestimmung des Slumpmaßes“
(Nebenseite 1. Ordnung zur Hauptseite „Versuche zur Konsistenzbestimmung“)

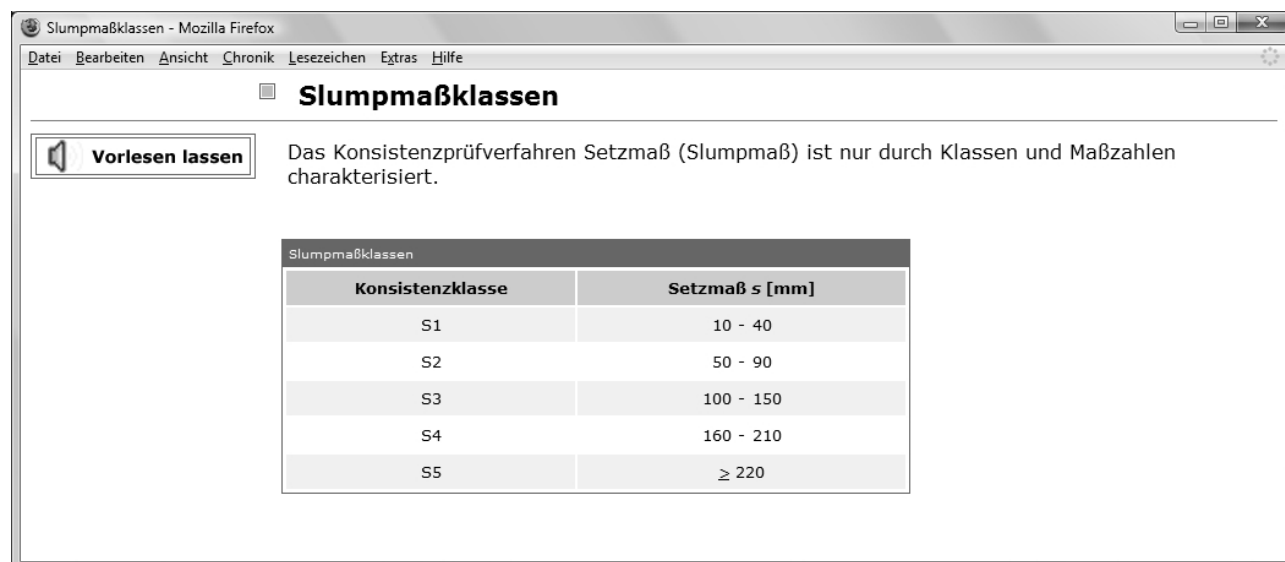


Abbildung G.8: Seite „Slumpmaßklassen“
(Nebenseite 2. Ordnung zur Nebenseite „Bestimmung des Slumpmaßes“)

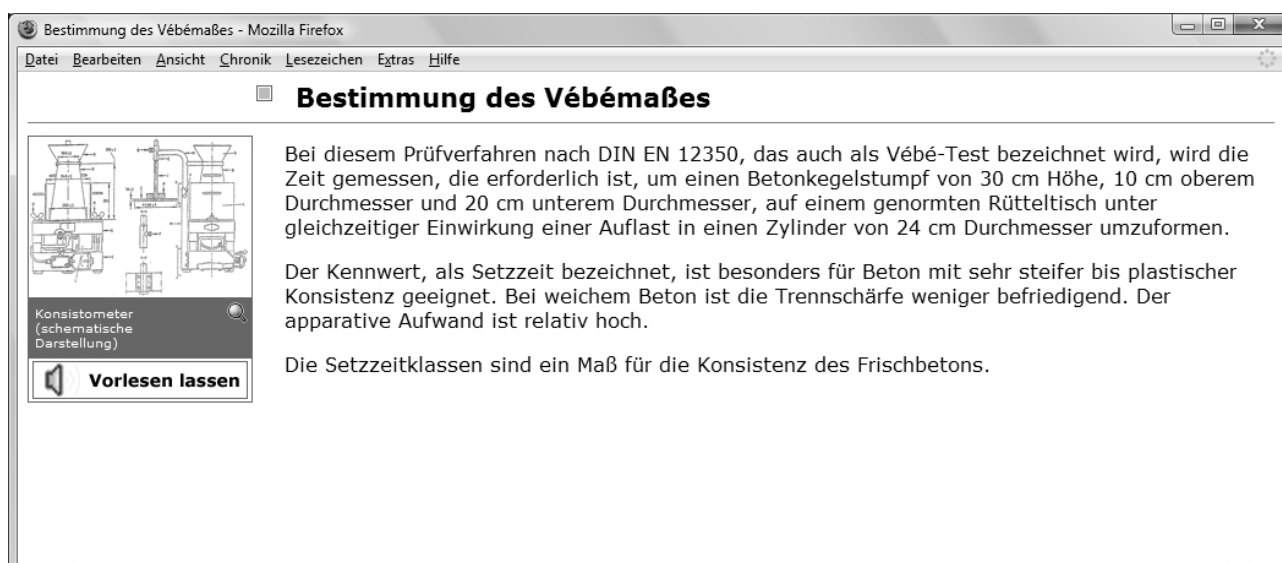


Abbildung G.9: Seite „Bestimmung des Vébémaßes“
(Nebenseite 1. Ordnung zur Hauptseite „Versuche zur Konsistenzbestimmung“)

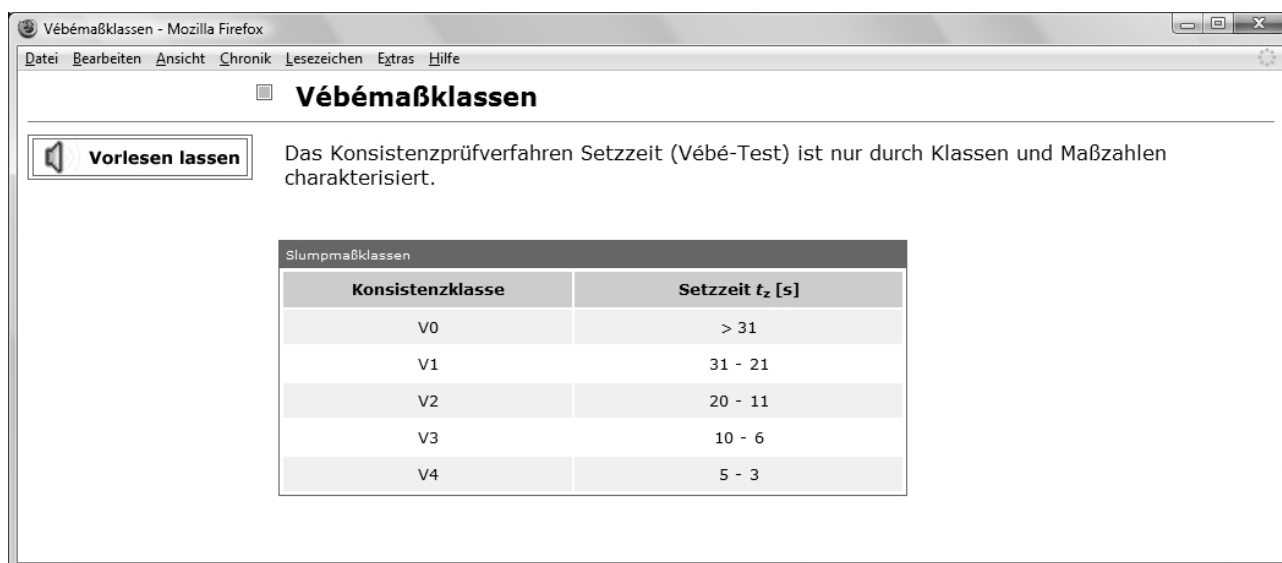


Abbildung G.10: Seite „Vébémaßklassen“
(Nebenseite 2. Ordnung zur Nebenseite „Bestimmung des Vébémaßes“)

Anhang H: Quelltext einer Wissensseinheit und ihre Anzeige im Browser

Im folgenden ist der Quelltext der WiBA-Net-Seite „Gefügearten des Zementits“ dargestellt. Die Inhaltsbereiche sind fett gedruckt. Zur besseren Lesbarkeit wurden Einrückungen und Zeilenumbrüche angepasst.

```
<html>
  <!-- #BeginTemplate "/Templates/nurlTextblock.dwt" -->
  <head>
    <!-- #BeginEditable "doctitle" -->
    <title>
      WiBA-Net >> Werkstoffe im Bauwesen und Architektur
    </title>
    <!-- #EndEditable -->
    <meta name="description" content="Das Lernnetzwerk f&uuml;r Werkstoffe im
    Bauwesen und Architektur.">
    <meta name="keywords" content="">
    <meta http-equiv="expires" content="0">
    <link rel="stylesheet" href="/Templates/css/pc-style.css" type="text/css">
    <script language="JavaScript" src="/Templates/scripts/wiba-netscript.js">
    </script>
  </head>
  <body>
    <a name="top"></a>
    <table width="100%" border="0" cellspacing="0" cellpadding="0">
      <tr>
        <td align="right" width="205">
          
        </td>
        <td class="titel">
          <!-- #BeginEditable "Überschrift" -->
          <B>
            Gef&uuml;gearten des Zementits
          </B>
          <!-- #EndEditable -->
        </td>
      </tr>
      <tr>
        <td colspan="2">
          <hr size="1" noshade>
        </td>
      </tr>
      <tr>
        <td rowspan="2">
          <!-- #BeginEditable "Extrafeld_vorlesen-lassen" -->
          <!-- #EndEditable -->
        </td>
        <td>
          <table border="0" cellpadding="5" cellspacing="0">
            <tr>
              <td class="textblock">
                <!-- #BeginEditable "Textblock" -->
                <p>
                  <B>
```

```

        Prim&auml;r;zementit
    </B>
    <BR>
    Prim&auml;r;zementit wird aus der Schmelze bei Kohlenstoffgehalten
    zwischen 4,3 M.-% bis 6,7 M.-% gebildet.
    <BR>
    <BR>
    <B>
        Sekund&auml;r;zementit
    </B>
    <BR>
    Bei der Umwandlung von Austenit in den kohlenstoff&auml;rmeren Ferrit
    wird Sekund&auml;r;zementit bevorzugt &uuml;ber die Korngrenzen
    ausgeschieden. Dabei entsteht um die Austenitk&ouml;rner
    Schalen&auml;r;zementit, der den Werkstoff verspr&ouml;nden l&auml;sst und ihn
    damit unbrauchbar macht (W&auml;rmebehandlung hilft!).
    <BR>
    <BR>
    <B>
        Terti&auml;r;zementit
    </B>
    <BR>
    Terti&auml;r;zementit wird bei untereutektoidem Stahl bei T = 723&deg;C
    &uuml;ber die Korngrenzen ausgeschieden. Dabei kommt es zur
    Verspr&ouml;dung von Gef&uuml;gebestandteilen. Sie sind nur in geringen
    Mengen vorhanden und im Schliffbild selten zu erkennen.
</p>
<P>
</P>
<!-- #EndEditable -->
</td>
</tr>
</table>
</td>
</tr>
<tr>
    <td class="copyright">
        <!-- #BeginEditable "copyright" -->
        <!-- #EndEditable -->
    </td>
</tr>
<tr>
    <td>
        
    </td>
    <td>
        
    </td>
</tr>
</table>
</body>
<!-- #EndTemplate -->
</html>

```

Im Webbrowser angezeigt sieht die Seite wie in der folgenden Abbildung dargestellt aus.

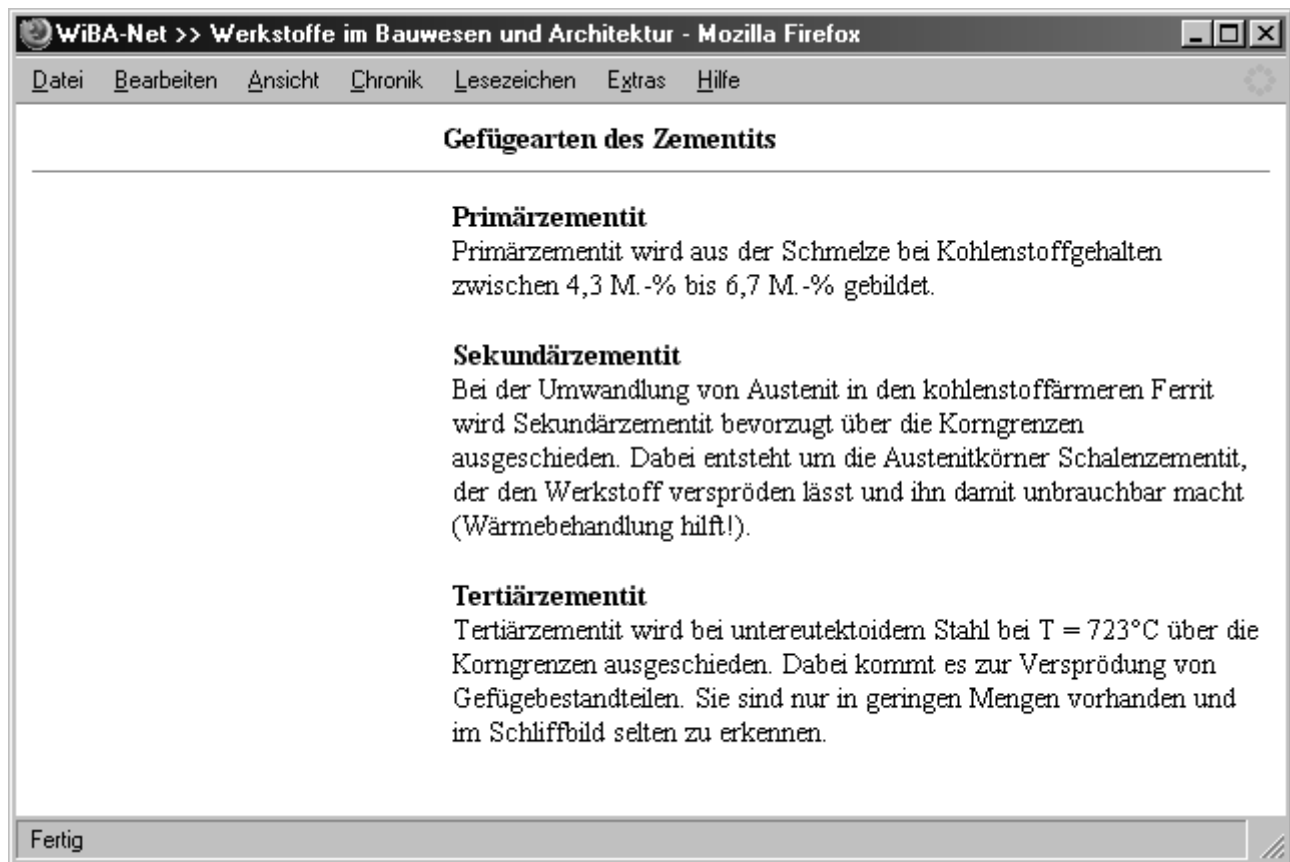


Abbildung H.1: Anzeige der WiBA-Net Seite "Gefügearten des Zementits" im Webbrowser Firefox

Anhang I: Seiten im automatisch generierten Lehrpfad

Im Folgenden sind die im Lehrpfad zur Suchanfrage „Gleichgewichtsfeuchte – Holz – Quellen“ enthaltenen Seiten dargestellt. Die Reihenfolge der Darstellung entspricht der Reihenfolge im Lehrpfad. Zu jeder Seite ist vermerkt, ob es sich um eine Haupt- oder Nebenseite handelt, bei Nebenseiten wird die zugehörige Hauptseite mit angegeben. Weiterführende Seiten werden nicht dargestellt. Eine Übersicht der Struktur des Lehrpfads kann Tabelle 6.1 entnommen werden.

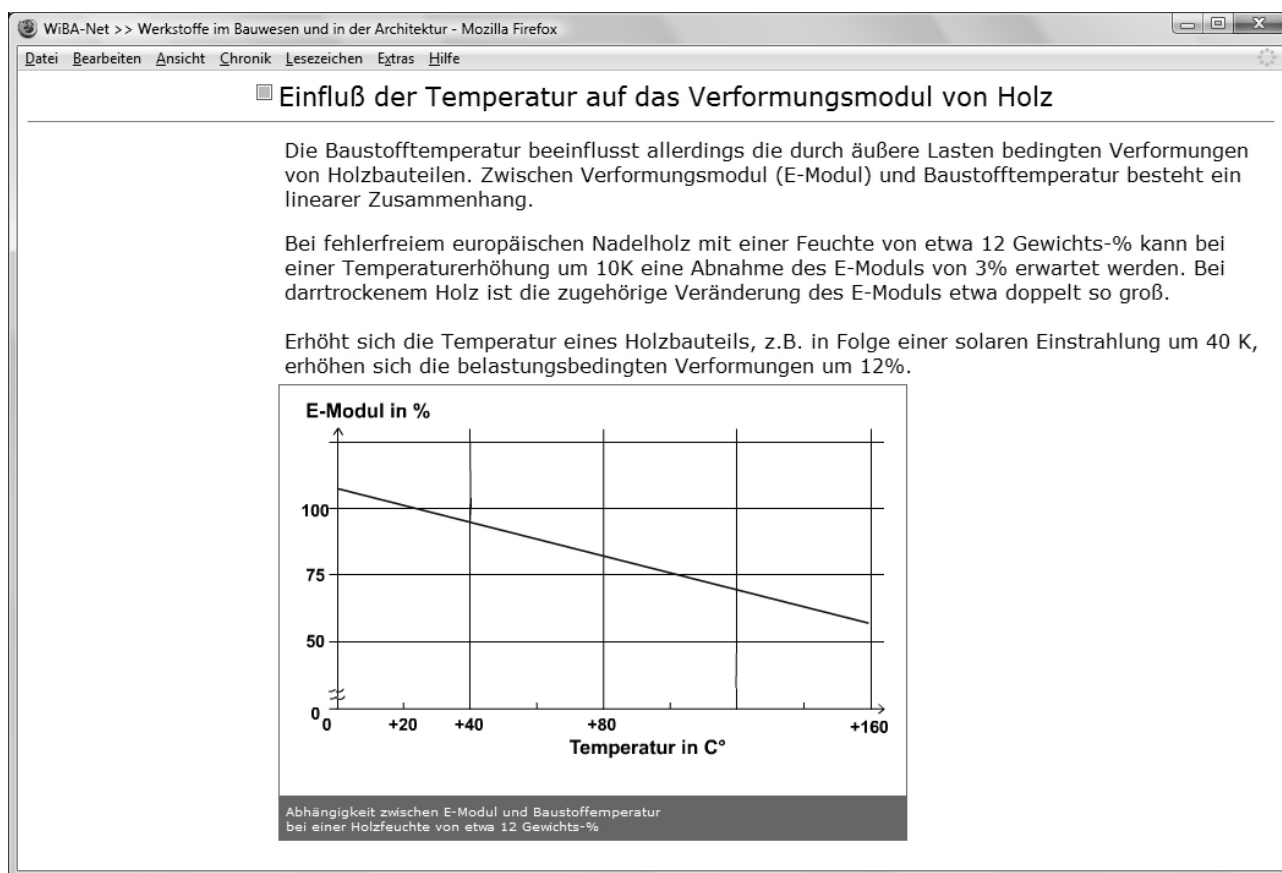


Abbildung I.1: Seite „Einfluss der Temperatur auf das Verformungsmodul von Holz“ (Hauptseite)

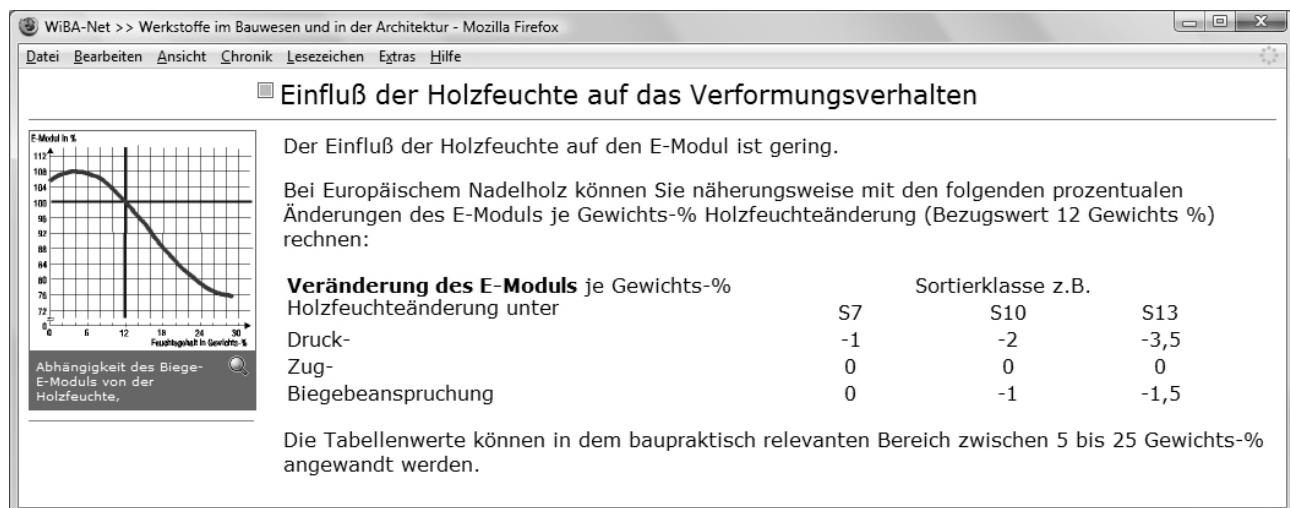


Abbildung I.2: Seite „Einfluss der Holzfeuchte auf das Verformungsverhalten“
(Nebenseite zur Hauptseite „Einfluss der Temperatur auf das Verformungsmodul von Holz“)

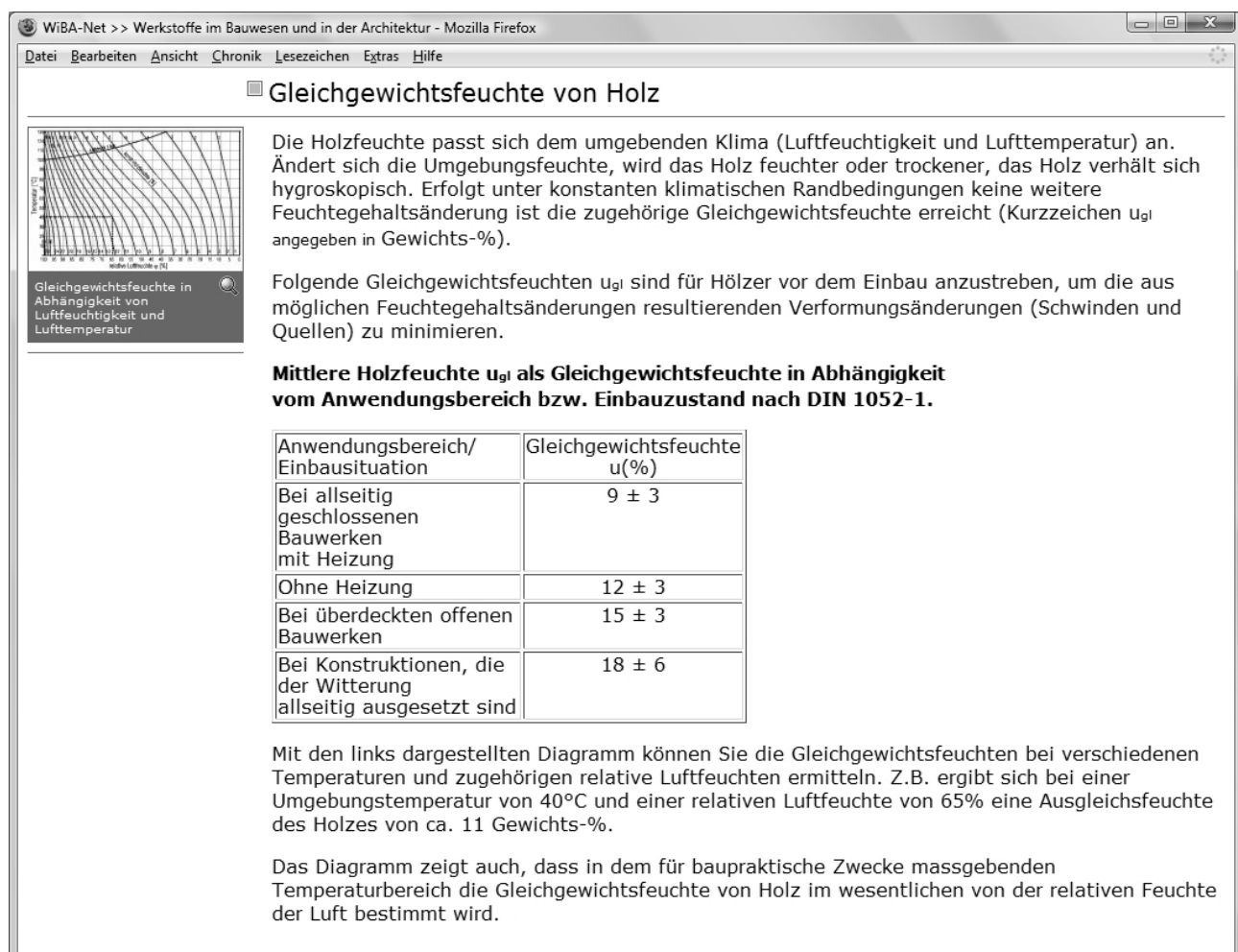


Abbildung I.3: Seite „Gleichgewichtsfeuchte von Holz“ (Hauptseite)

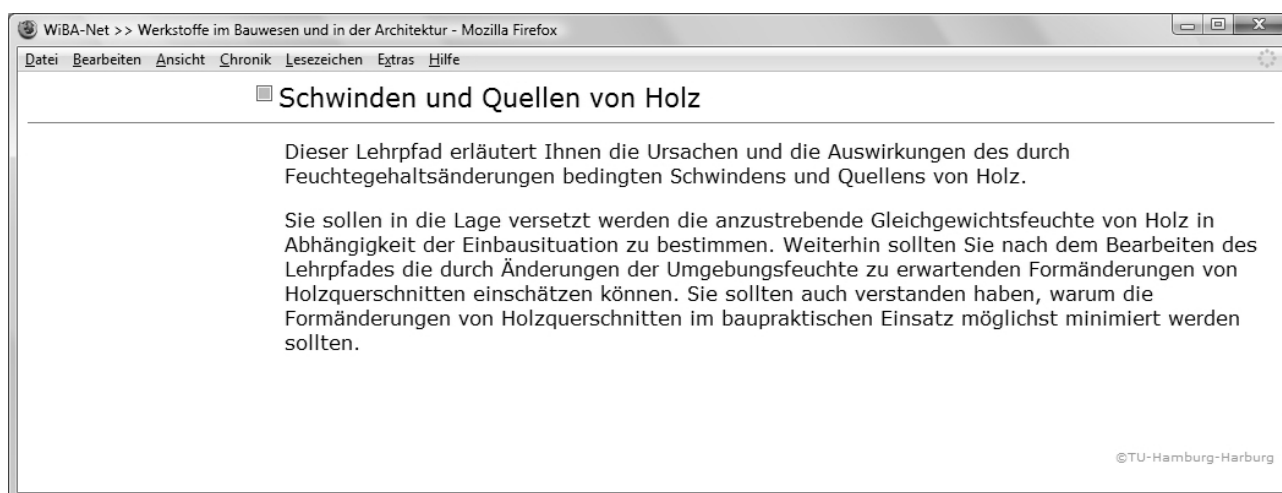


Abbildung I.4: „Schwinden und Quellen von Holz“
(Nebenseite zur Hauptseite „Gleichgewichtsfeuchte von Holz“)

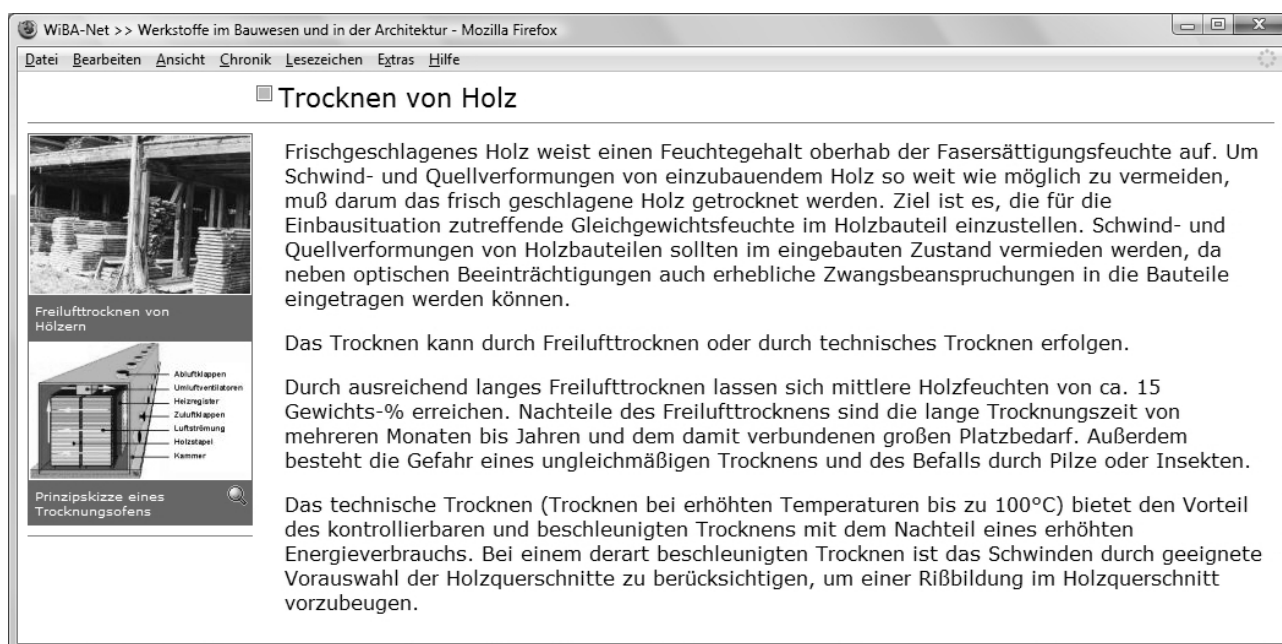
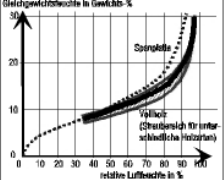


Abbildung I.5: Seite „Trocknen von Holz“ (Hauptseite)

WiBA-Net >> Werkstoffe im Bauwesen und in der Architektur - Mozilla Firefox

Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe

Sorptionsisotherme und Mechanismen der Feuchteaufnahme



Die Gleichgewichtsfeuchte kann für eine definierte Temperatur in Abhängigkeit der das Bauteil umgebenden Luftfeuchte dargestellt werden. Diese Darstellung wird als Sorptionsisotherme bezeichnet. Da, wie bereits festgestellt, für den baupraktisch relevanten Temperaturbereich die Gleichgewichtsfeuchte nahezu unabhängig von der Temperatur ist, können Sie näherungsweise eine einzige temperaturunabhängige Sorptionsisotherme verwenden.

Die Aufnahme von Wasser erfolgt bis ca. 8 Gewichts-% Holzfeuchte durch **Chemosorption**. Die Chemosorption ist eine Wechselwirkung zwischen den OH-Gruppen der Zellulose und den Wasserdipolen.

Weiteres Wasser kann durch **Adsorption** an den Faseroberflächen gebunden werden. Dieser Prozess, der auf der Wasserstoffbrückenbildung basiert, ist nach der Aufnahme von ca. 7 weiteren Gewichts-% (Holzfeuchte 15 Gewichts-%) abgeschlossen.

Eine Wasseraufnahme über 15 Gewichts-% findet durch **Kapillarkondensation** in den Fasern statt, und zwar solange bis die **Fasersättigungsfeuchte** erreicht wird. Die Fasersättigungsfeuchte ist abhängig vom dem zu Verfügung stehenden Volumen der Kapillaren und damit von der Holzart. Die Fasersättigungsfeuchte liegt zwischen 20 und 35 Gewichts-%.

Eine weitere Wasseraufnahme über die Fasersättigungsfeuchte hinaus kann nur noch durch kapillares Aufsaugen erfolgen.

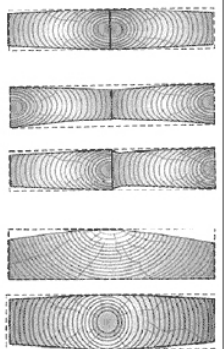
©TU-Hamburg-Harburg

Abbildung I.6: Seite „Sorptionsisotherme und Mechanismen der Feuchteaufnahme“
(Nebenseite zur Hauptseite „Trocknen von Holz“)

WiBA-Net >> Werkstoffe im Bauwesen und in der Architektur - Mozilla Firefox

Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe

Charakteristische Eigenschaften - Schwinden und Quellen



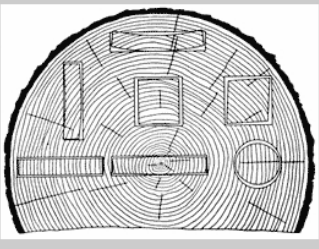
Holz arbeitet

Darunter ist das **Schwinden und Quellen** sowie das „Werfen des Holzes“ zu verstehen. Bei der Aufnahme von Feuchtigkeit durch chemisch-physikalische Vorgänge vergrößert sich die Zellstruktur, bei der Abgabe verkleinert sie sich. Auch nach der Verarbeitung des Holzes zu Balken, Brettern u.s.w. bleibt die Eigenschaft des Schwindens und Quellens bestehen. Die geringste Feuchte-Ausdehnung ist längs zur Faser gegeben (günstigstes Feuchteverhalten).

Quellmaße bei einer Änderung der Holzfeuchte um 1% (durchschnittlich, unterhalb der Fasersättigung):

radial	0,12%
tangential	0,24%
längs	0,01%

Verzerrung von Holzquerschnitten infolge des Schwindens. Die Holzstruktur im Bereich des Kernholzes schwindet (und quillt) nicht so stark wie beim Splintholz, da das Feuchteabgabe- (bzw. Feuchteaufnahme-) Vermögen dort geringer ist. Dieses Verhalten erklärt die Krümmung der Querschnitte wie im unteren Bild dargestellt.



Schwindverhalten bei verleimten Brettquerschnitten, günstig ist eine gespiegelte Verleimung kernseitig

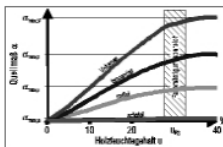
KG Universität Leipzig

Abbildung I.7: Seite „Charakteristische Eigenschaften – Schwinden und Quellen“ (Hauptseite; der graue Hintergrund weist darauf hin, dass es sich um eine Seite aus dem Bereich der Architektur handelt)

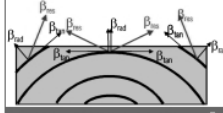
WiBA-Net >> Werkstoffe im Bauwesen und in der Architektur - Mozilla Firefox

Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe

■ Quell- und Schwindmaße von Holz



Quellkurven in unterschiedliche Richtungen



Vektorielle Darstellung der Schwindverformung

Mögliche Änderungen des Feuchtegehaltes von Holz unterhalb der Fasersättigung führen zu Formänderungen. Diese als Quellen und Schwinden bezeichneten Formänderungen sind in den verschiedenen Richtungen unterschiedlich groß. Das Schwind- bzw. Quellmaß bewegt sich für europäisches Bauholz in den folgenden Größenordnungen (angegeben sind maximale Maße, ausgehend von der Fasersättigungsfeuchte):

achsial	radial	tangential	volumenmäßig
0,4%	4,3%	8,2%	12%
bzw. allgemein im Verhältnis von ca.			
1	10	20	30

Feuchtegehaltsänderungen oberhalb des Fasersättigungspunktes führen nicht zu nennenswerten Quell- bzw. Schwindverformungen.

Die Quell- und Schwindprozesse können zu Rissen im Holzquerschnitt führen.

TU-Hamburg-Harburg

Abbildung I.8: Seite „Quell- und Schwindmaße von Holz“

(Nebenseite zur Hauptseite „Charakteristische Eigenschaften – Schwinden und Quellen“)

WiBA-Net >> Werkstoffe im Bauwesen und in der Architektur - Mozilla Firefox

Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe

■ Relevante Eigenschaften der für Bauzwecke eingesetzten Holzarten



Die Kiefer ist ein typischer Vertreter der im Bauwesen eingesetzten Nadelholzarten.



Als Vertreter der Laubholzarten ist in diesem Bild die Eiche gezeigt.

Es gibt ca. 25000 Holzarten auf der Erde, von denen ca. nur 2000 Arten klassifiziert sind. Davon werden ungefähr 300 Arten gewerblich genutzt. Die Hauptunterteilung erfolgt in Laub- und Nadelholzarten.

Die für ein Bauvorhaben wesentlichen Eigenschaften von Hölzern sind in der Regel

- die **Dichte**,
- die **Festigkeit**,
- die **Beständigkeit**
 - bei Bewitterung,
 - gegen einen Pilzbefall,
 - oder gegen einen Insektenbefall,
- und das **Stehvermögen**.

Das **Stehvermögen** beschreibt das Formänderungsverhalten von verarbeitetem Holz bei wechselnden Umgebungsfeuchten. Neben den Absolutwerten der Quell- und Schwindmaße ist auch die Formhaltigkeit des Holzquerschnittes bei der Beurteilung des Stehvermögens zu beachten. Die Formhaltigkeit eines Holzquerschnittes wird durch das Verhältnis der radialen und tangentialen Schwindverformungen und somit auch durch den Einschnitt des Holzes (Lage des Querschnitts im Stamm) bestimmt.

Diese für Bauhölzer wesentlichen Eigenschaften, anhand derer die Auswahl für den vorgesehenen Einsatzzweck erfolgt, werden im Folgenden für einige ausgewählte Holzarten dargestellt.

Die Angaben erfolgen, falls dies der richtungsabhängige Aufbau von Holz erfordert, für die zu unterscheidenden Richtungen parallel und senkrecht zur Faser.

Neben den einheimischen, europäischen Holzarten werden auch die Eigenschaften einiger überseeischer Holzarten betrachtet.

Abbildung I.9: Seite „Relevante Eigenschaften der für Bauzwecke eingesetzten Holzarten“

(Nebenseite zur Hauptseite „Charakteristische Eigenschaften – Schwinden und Quellen“)

WiBA-Net >> Werkstoffe im Bauwesen und in der Architektur - Mozilla Firefox

Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe

■ Schwindverformungen

Schwindrichtungen von Schnittholzern, die aus unterschiedlichen Querschnittsbereichen entnommen wurden.

Durch die unterschiedlichen Schwindmaße in radiale und tangential Richtung im Verhältnis $10:20=1:2$ verformt sich Holz beim Trocknen ungleichmäßig. Die Verformungen sind davon abhängig, aus welchem Teil des Querschnittes das Holz entnommen wurde.

Das **Stehvermögen** beschreibt das Formänderungsverhalten von verarbeitetem Holz bei wechselnden Umgebungsfeuchten. Neben den Absolutwerten der Quell- und Schwindmaße ist auch die Formhaltigkeit des Holzquerschnittes bei der Beurteilung des Stehvermögens zu beachten. Die Formhaltigkeit eines Holzquerschnittes wird durch das Verhältnis der radialen und tangentialen Schwindverformungen und somit auch durch den Einschnitt des Holzes (Lage des Querschnitts im Stamm) bestimmt.

Dieses können Sie mit der folgenden Animation ausprobieren. Versuchen Sie die nach dem Schwindvorgang vorliegende Querschnittsform vorherzusagen.

Zur Darstellung der unterschiedlichen Schwindverformungen in Abhängigkeit der Querschnittslage bitte mit der Maus auf die gekennzeichneten Bereiche fahren und die Maus nicht bewegen.

TU-Hamburg-Harburg

Abbildung I.10: Seite „Schwindverformungen“

(Nebenseite zur Hauptseite „Charakteristische Eigenschaften – Schwinden und Quellen“)

Die folgenden Seiten wurden nachträglich aus dem Lehrpfad ausgeschlossen:

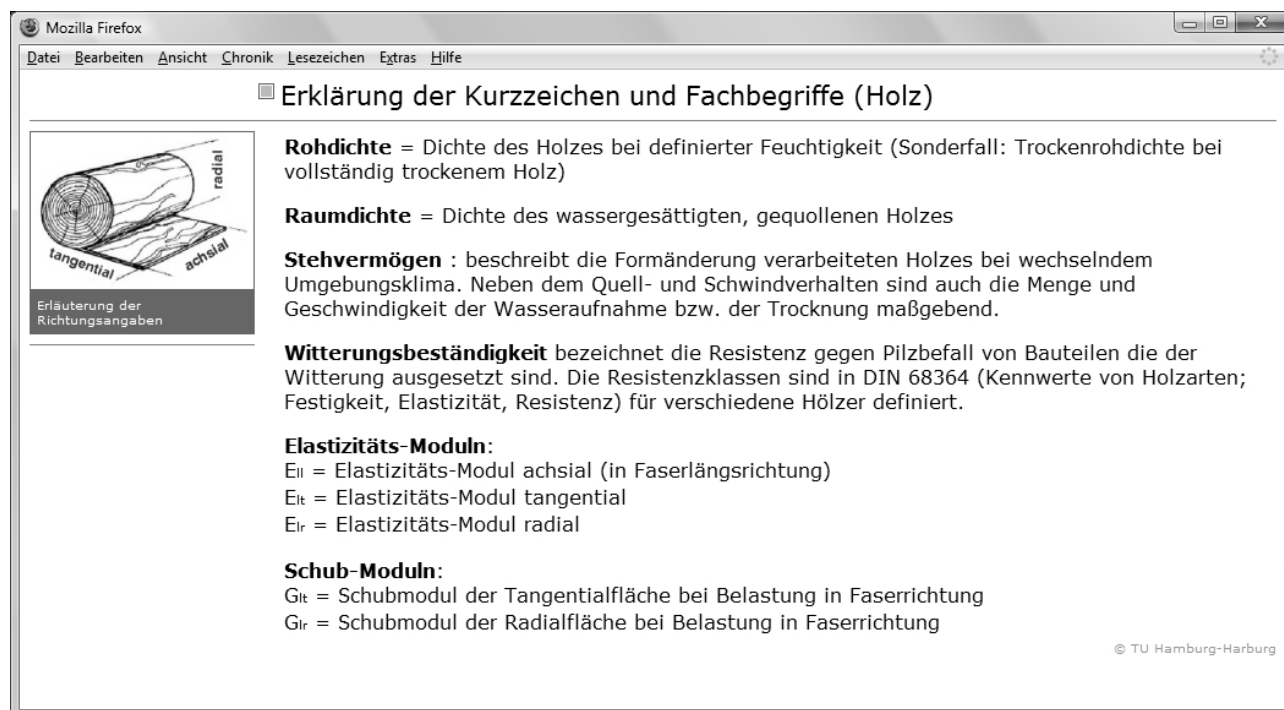


Abbildung I.11: Seite „Erklärung der Kurzzeichen und Fachbegriffe (Holz)“
(aus dem Lehrpfad entfernte Seiten)

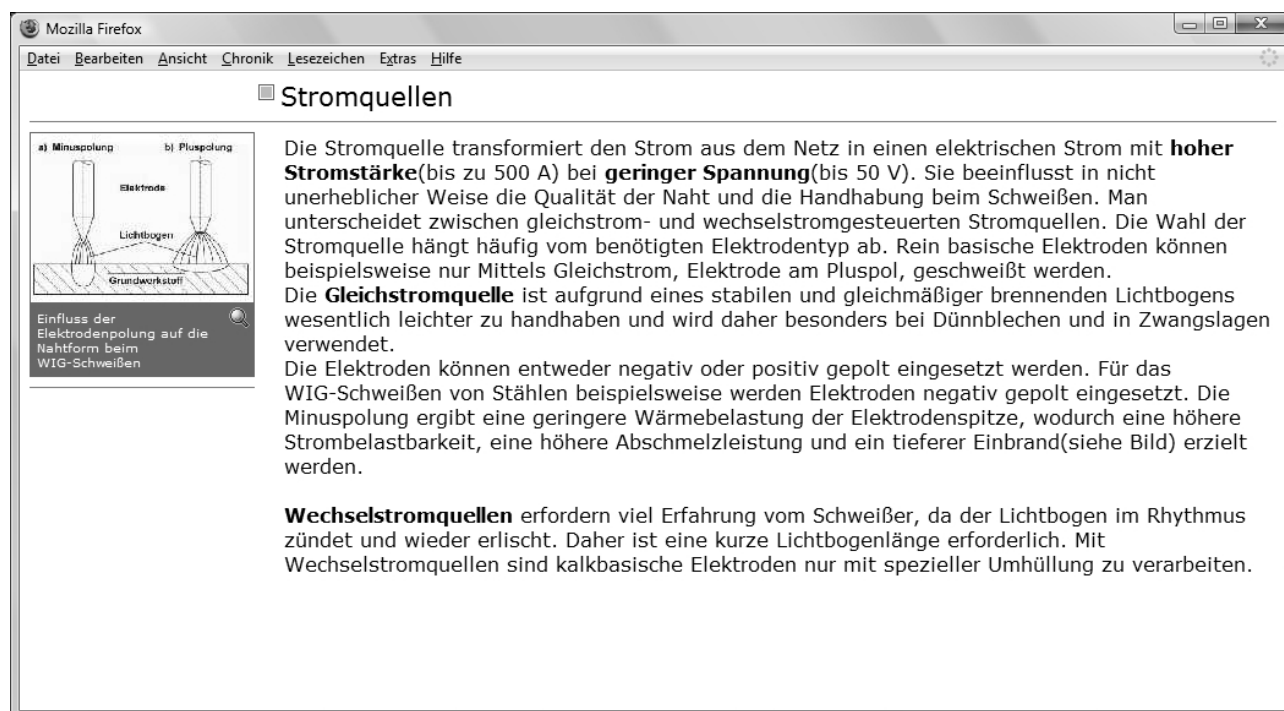


Abbildung I.12: Seite „Stromquellen“
(aus dem Lehrpfad entfernte Seiten)

Anhang J: Ausgewertete Lehrpfade

Im Folgenden sind mit dem vorgeschlagenen System erstellte Lehrpfade zu 62 Suchanfragen dargestellt. Die Lehrpfade sind alphabetisch nach den Suchanfragen geordnet. Die Wissensseinheiten in einem Lehrpfad sind nach der für sie mit dem System berechneten Relevanz R geordnet.

Die in den Lehrpfaden enthaltenen Wissensseinheiten wurden auf ihre Relevanz bezüglich des durch die jeweiligen Suchbegriffe implizierten Themas untersucht. Dabei wurde zum einen die Relevanz bezüglich des die Suchbegriffe eng umfassenden Kontextes und der durch die Kombinationen einzelner Suchbegriffe aufgespannten Querschnittsthemen betrachtet. Der jeweilige Relevanzwert ist in der Spalte R1 verzeichnet. Zum anderen wurde eine auf ein weiter gefasstes Themengebiet, das auch mit dem Hauptthema lediglich themenverwandte Wissensseinheiten beinhaltet, bezogene Relevanz untersucht. Diese Werte sind in der Spalte R2 dargestellt. In den Spalten R1 und R2 bezeichnet eine „1“ jeweils eine relevante Seite und eine „0“ eine nicht relevante Seite. Die jeweilige Summe ist unterhalb der Spalte angegeben. Die auf die Gesamtanzahl der Seiten im Lehrpfad bezogene Anzahl $\Sigma R1$ bzw. $\Sigma R2$ der jeweils relevanten Seiten ergibt die Precision P nach der eng gefassten Definition bzw. die Precision P' nach der weit gefassten Definition als prozentualen Wert.

Wie in Fußnote 50 in Abschnitt 5.1.3.2 beschrieben, sind Wissensseinheiten am Ende eines Lehrpfads, die keine Relevanz für den entsprechenden Lehrpfad aufweisen und deren Relevanzwert R einen Abstand von mehr als 0,1 zum nächsthöheren Relevanzwert hat, unterstrichen dargestellt.

Suchbegriffe: Abbinden – Zement Anzahl Seiten: 12 P: 50,00% P': 58,33%				57,94	Physikalisch abbindende Klebstoffe	0	0
						Σ:	6 7
Suchbegriffe: Abbindevorgang – Estrich – Klebstoff Anzahl Seiten: 7 P: 85,71% P': 100,00%							
R	Titel	R1	R2	R	Titel	R1	R2
80,86	Hauptbestandteile des Zement	1	1	73,41	Abbindevorgänge beim Kleben	1	1
80,49	Hauptarten des Zements	0	1	66,88	Estrich auf Trennlagen	0	1
80,31	Hauptbestandteile	1	1	60,16	Reaktionsklebstoffe, Einkomponentenkleber	1	1
79,58	Erhärten des Betons	1	1	49,96	Physikalisch abbindende Klebstoffe	1	1
75,44	Zement	1	1	42,81	Magnesiaestrich	1	1
70,57	Portlandzementklinker (K)	1	1	41,48	Anhydritestrich	1	1
69,32	Systematik der Klebstoffe in Anlehnung an DIN 16 920	0	0	40,11	Reaktionsklebstoffe - Zweikomponentenkleber	1	1
67,72	Das Verfahren der Zementherstellung	0	0			Σ:	6 7
66,57	Hydratationswärme	1	1				
63,83	Verfahren der Zementherstellung	0	0				
63,21	Abbindevorgänge beim Kleben	0	0				

Suchbegriffe: Aluminium – Verwendung**Anzahl Seiten:** 11**P:** 81,82%**P':** 90,91%

R	Titel	R1	R2
71,63	Gewinnung	1	1
70,93	Einsatzgebiete von Aluminium	1	1
66,61	Aluminium-Legierungen	1	1
57,13	Aluminiumprodukte im Bauwesen	1	1
56,33	Kupfer-Legierungen	0	0
55,32	Rohstoffe	1	1
51,56	Angriff durch nichtmetallisch-anorganische Baustoffe	1	1
47,93	Ausgangsmaterial	1	1
46,40	Übersicht wichtiger Nichteisenmetalle (1)	1	1
45,06	Verwendungsgeschichte von Eisen und Stahl	0	1
39,69	Formbarkeit von NE-Metallen	1	1
Σ:		9	10

Suchbegriffe: Anwendung – Laubholz – Nadelholz**Anzahl Seiten:** 9**P:** 77,78%**P':** 100,00%

R	Titel	R1	R2
79,65	Nadelhölzer	1	1
78,92	Überblick über die im Bauwesen verwendeten Holzarten	1	1
75,55	Festigkeitsklassen von Bauholz für tragende Zwecke	1	1
73,09	Steifigkeitseigenschaften der Festigkeitsklassen von Bauholz für tragende Zwecke	1	1
71,51	Relevante Eigenschaften der für Bauzwecke eingesetzten Holzarten	1	1
67,44	Laubholz	1	1
48,57	Zusammenstellung der technischen Kennwerte von Laubhölzern	1	1
43,12	Aufbau von Holz - Fette und Harze	0	1
29,60	Zeitschiene der Entwicklung des Holzbaus	0	1
Σ:		7	9

Suchbegriffe: Asphalt – Bitumen**Anzahl Seiten:** 11**P:** 72,73%**P':** 90,91%

R	Titel	R1	R2
80,28	Klassifizieren von Bitumen	0	1
80,21	Asphalt mit Bindemittelüberschuß - Gußasphalt	1	1
73,08	Hartbitumen	1	1
72,25	Gussasphaltestrich	1	1
70,29	Elastomer- und Plastomerbitumen	1	1
70,17	Asphalt - Ausgangsstoffe für die Mischung	1	1
68,61	Oxidationsbitumen	1	1
65,99	Aufbau einer Bitumenbahn	0	1
65,88	Bitumen - Verarbeitungsformen	1	1
65,31	Anforderungen an die Zusammensetzung von "Walz"-Asphalt	1	1
47,31	<u>Estriche - Materialien</u>	0	0
Σ:		8	10

Suchbegriffe: Bestandteile – Beton**Anzahl Seiten:** 15**P:** 60,00%**P':** 93,33%

R	Titel	R1	R2
89,17	Hauptbestandteile des Zement	0	1
85,04	Hauptbestandteile	0	1
79,31	Zement	1	1
79,12	Zuschlag / Schädliche Bestandteile	1	1
73,68	Nebenbestandteile des Zement	0	1
73,42	Hauptarten des Zements	0	1
70,88	Ausgangsstoffe des Betons (Zementbeton)	1	1
70,69	Betonzusatzstoffe / Kennwerte	1	1
69,64	Betonzusatzstoffe / inaktive Stoffe	1	1
69,37	Wirkungsgruppen und Kennzeichnung der Betonzusatzmittel	1	1
68,38	Hydrationsbeeinflussende Stoffe	0	1
68,29	Betonzusatzmittel / Stabilisierer (ST)	1	1
68,22	Betonzusatzstoffe / Einführung	1	1
67,62	Betonzusatzstoffe / anorganisch-mineralische Zusatzstoffe	1	1
60,00	Endkriechzahl für Normalbeton und	0	0

feuchte Umgebungsbedingungen			
	Σ:	9	14

Suchbegriffe: Bestandteil – Zement

Anzahl Seiten: 14

P: 78,57%

P': 100,00%

R	Titel	R1	R2
89,17	Hauptbestandteile des Zement	1	1
88,08	Nebenbestandteile des Zement	1	1
80,27	Hauptbestandteile	1	1
79,24	Erhärten des Betons	0	1
78,49	Hauptarten des Zements	1	1
78,02	Zement	1	1
77,21	Das Verfahren der Zementherstellung	1	1
75,96	Verfahren der Zementherstellung	1	1
74,55	Portlandzementklinker (K)	1	1
74,41	Hydratationsbeeinflussende Stoffe	1	1
66,73	Zusammensetzung von Zement	1	1
65,74	Puzzolane	0	1
65,24	Ausgangsstoffe des Betons (Zementbeton)	0	1
65,10	Mahlen des Zements	1	1
	Σ:	11	14

Suchbegriffe: Beton – Ausgangsstoff

Anzahl Seiten: 16

P: 68,75%

P': 75,00%

R	Titel	R1	R2
89,17	Ausgangsstoffe des Betons (Zementbeton)	1	1
85,04	Ausgangsstoffe	1	1
76,07	Betonzusatzstoffe / Latent-hydraulische Stoffe	1	1
71,01	Betonzusatzstoffe / inaktive Stoffe	1	1
70,51	Betonzusatzstoffe / anorganisch-mineralische Zusatzstoffe	1	1
69,86	Wirkungsgruppen und Kennzeichnung der Betonzusatzmittel	1	1
66,55	Faserbeton / Stahlfaserarten	0	0
65,55	Betonzusatzstoffe / Puzzolanische Stoffe	1	1
64,26	Exkurs Stahlbeton: Volumenvergrößerung	0	0

64,03	Beton	1	1
63,64	Betonzusatzmittel / Einleitung	1	1
63,57	Betonstähle	1	1
63,26	Frischbetoneigenschaften	0	1
62,97	Betonzusatzstoffe / Einführung	1	1
60,00	Endkriechzahl für Normalbeton und feuchte Umgebungsbedingungen	0	0
55,60	Endkriechzahl für Normalbeton und trockene Umgebungsbedingungen	0	0
	Σ:	11	12

Suchbegriffe: Beton – Bewehrung – Korrosion

Anzahl Seiten: 15

P: 66,67%

P': 66,67%

R	Titel	R1	R2
88,00	Expositionsklassen hinsichtlich Bewehrungskorrosion	1	1
84,62	Korrosion der Bewehrung	1	1
81,23	Dauerhaftigkeit von bewehrten Betonkonstruktionen	1	1
79,96	Expositionsklasse / X0	1	1
77,18	Exkurs Stahlbeton: Depassivierung	1	1
76,86	Korrosion der Oberfläche (Betonkorrosion) des Bauteils	1	1
76,41	Wirkungsgruppen und Kennzeichnung der Betonzusatzmittel	0	0
72,65	Faserbeton / Stahlfaserarten	0	0
71,72	Allgemeines zum Bezeichnungssystem für Betonstahl	1	1
71,44	Betonstähle	1	1
67,79	Zusatzstoffe	0	0
67,07	Betonzusatzmittel / Einleitung	0	0
66,82	Allgemeines zum Betonstahl	1	1
63,08	Zusatzmittel	0	0
43,45	Expositionsklassen / Einteilung	1	1
	Σ:	10	10

Suchbegriffe: Beton – Bewehrung – Korrosion – Stahl

Anzahl Seiten: 23

P: 56,52%

P': 56,52%

R	Titel	R1	R2
84,43	Allgemeines zum Bezeichnungssystem für Betonstahl	1	1

79,90	Faserbeton / Stahlfaserarten	0	0	79,89	Betonstähle	1	1
75,42	Dauerhaftigkeit von bewehrten Betonkonstruktionen	1	1	77,76	Brucheinschnürung (Z)	1	1
74,87	Expositionsklasse / X0	1	1	77,24	Spannungs-Dehnungs-Diagramm eines kaltverformten Stahls	1	1
73,61	Definitionen zu Betonstahl	1	1	75,03	Allgemeines zum Betonstahl	1	1
72,25	Betonstähle	1	1	73,25	Definitionen zu Betonstahl	0	0
71,68	Expositionsklassen hinsichtlich Bewehrungskorrosion	1	1	72,89	Bruchdehnung (A)	1	1
71,15	Exkurs Stahlbeton: Depassivierung	1	1	71,80	Spannungs-Dehnungs-Diagramm eines warmverformten Stahls	1	1
69,13	Beispiele zum Bezeichnungssystem von Betonstahlmatten	0	0	70,61	Eigenschaften von Betonstahl	1	1
68,30	Plötzliches Versagen eines Stahlbetonbauteils mit zu geringer Mindestbewehrung	0	0	70,51	Gleichmaßdehnung (Ag) beim kaltverformten Stahl	0	0
67,93	Allgemeines zu Betonstahlmatten	0	0	69,65	Wirkungsgruppen und Kennzeichnung der Betonzusatzmittel	0	0
67,31	Allgemeines zum Betonstahl	1	1	69,14	Einschnürungsbereich beim kaltverformten Stahl	0	0
67,14	Korrosion der Oberfläche (Betonkorrosion) des Bauteils	1	1	68,91	Faserbeton / Stahlfaserarten	0	0
65,11	Wirkungsgruppen und Kennzeichnung der Betonzusatzmittel	0	0	51,57	Makroskopisches Bruchverhalten	1	1
64,79	Beispiel für die Bezeichnung von Bewehrungsdraht	0	0	Σ: 10 12			
64,63	Korrosion der Bewehrung	1	1	Suchbegriffe: Beton – Brücke			
63,42	Zusatzstoffe	0	0	Anzahl Seiten: 13			
62,43	Betonzusatzstoffe / Einführung	1	1	P: 15,38%			
59,08	Betonzusatzmittel / Einleitung	0	0	P': 46,15%			
55,72	Carbonatisierung / Schadensvermeidung	1	1	R	Titel	R1	R2
55,67	Zusatzmittel	0	0	75,84	Wirkungsgruppen und Kennzeichnung der Betonzusatzmittel	0	0
55,44	Beispiel für die Bezeichnung von Betonstabstahl	0	0	75,68	Form und Funktion - Tragwerke aus Beton (2)	1	1
47,82	Faserbeton / Anwendungsgebiete	1	1	74,97	Historische Holzbrücken (1)	0	0
Σ: 13 13				72,83	Betonzusatzstoffe / inaktive Stoffe	0	0
Suchbegriffe: Beton – Bruch – Stahl Anzahl Seiten: 17 P: 58,82% P': 70,59%				69,50	Schrägseilbrücken	0	1
				68,80	Historische Holzbrücken (2)	0	0
				67,23	Betonstähle	0	0
				66,54	Ausgangsstoffe des Betons (Zementbeton)	0	1
				66,41	Zur Konstruktion der Schrägseilbrücken	0	1
				66,15	Bogenbrücken	1	1
				65,99	Betonzusatzstoffe / anorganisch-mineralische Zusatzstoffe	0	0
				63,95	Frischbetoneigenschaften	0	0
				52,60	Düsseldorfer Brückenfamilie von Fritz Leonhardt	0	1
				Σ: 2 6			

Suchbegriffe: Beton - Druckfestigkeit				
Anzahl Seiten: 18				
P: 72,22%				
P': 88,89%				
R	Titel	R1	R2	
89,17	Einflüsse auf die Druckfestigkeit von Beton	1	1	
89,17	Druckfestigkeitsklassen für Normal- und Schwerbeton	1	1	
89,17	Druckfestigkeitsklassen für Leichtbeton	1	1	
89,17	Betondruckfestigkeitsprüfung / Filme	1	1	
89,17	Betondruckfestigkeitsprüfung	1	1	
89,17	Betondruckfestigkeit / Prüfkörpergestalt und -abmessungen	1	1	
79,44	Betonfestigkeitsklassen	1	1	
75,56	Druckfestigkeit	1	1	
72,94	Spannungs-Dehnungs-Linie von Beton	1	1	
72,94	Festigkeitsklassen und Rohdichte	1	1	
70,87	Betonzusatzstoffe / Einführung	0	1	
70,82	Betonzusatzmittel / Einleitung	0	1	
69,88	Ausgangsstoffe des Betons (Zementbeton)	0	1	
69,66	Betonzusatzstoffe / inaktive Stoffe	0	0	
69,56	Eigenschaften von Mauersteinen aus Normal- und Leichtbeton	1	1	
69,37	Zusammenhang zwischen Zug- und Druckfestigkeit	1	1	
69,23	Betonzusatzstoffe / anorganisch-mineralische Zusatzstoffe	1	1	
67,43	Wirkungsgruppen und Kennzeichnung der Betonzusatzmittel	0	0	
Σ:		13	16	

Suchbegriffe: Beton – Druckfestigkeit – Prüfung				
Anzahl Seiten: 10				
P: 90,00%				
P': 90,00%				
R	Titel	R1	R2	
89,17	Einflüsse auf die Druckfestigkeit von Beton	1	1	
89,17	Druckfestigkeitsklassen für Normal- und Schwerbeton	1	1	
89,17	Druckfestigkeitsklassen für Leichtbeton	1	1	
89,17	Betondruckfestigkeitsprüfung / Filme	1	1	
89,17	Betondruckfestigkeit / Prüfkörpergestalt und -abmessungen	1	1	

74,12	Festigkeitsklassen und Rohdichte	1	1
73,84	Der Einfluss des w/z-Werts auf die Betondruckfestigkeit	1	1
71,10	Betonfestigkeitsklassen	1	1
62,21	Druckfestigkeit	1	1
58,80	Eigenschaften von Mauersteinen aus Normal- und Leichtbeton	0	0
Σ:		9	9

Suchbegriffe: Beton – Druckfestigkeit – Zugfestigkeit				
Anzahl Seiten: 22				
P: 68,18%				
P': 77,27%				

R	Titel	R1	R2
89,17	Druckfestigkeitsklassen für Normal- und Schwerbeton	1	1
89,17	Druckfestigkeitsklassen für Leichtbeton	1	1
89,17	Betondruckfestigkeitsprüfung / Filme	1	1
89,17	Betondruckfestigkeit / Prüfkörpergestalt und -abmessungen	1	1
84,90	Einflüsse auf die Druckfestigkeit von Beton	1	1
74,04	Biegezugfestigkeit	1	1
73,68	Faserbeton / Verhalten bei Zug- und Biegebeanspruchung	1	1
72,96	Spannungs-Dehnungs-Linie von Beton	1	1
69,67	Faserbeton / Stahlfaserarten	0	1
69,26	Zentrische Zugfestigkeit	1	1
67,73	Faserbeton / Verhalten bei Druckbeanspruchung	1	1
66,25	Betonfestigkeitsklassen	1	1
65,27	Druckfestigkeit	1	1
59,90	Zug-, Biege-, Druck- und Schubfestigkeit	0	0
59,34	Spannungs-Dehnungs-Diagramm eines warmverformten Stahls	0	0
55,45	Zusammenhang zwischen Zug- und Druckfestigkeit	0	1
48,78	Festigkeitsklassen und Rohdichte	1	1
44,79	Festbeton	1	1
44,41	Der Einfluss des w/z-Werts auf die Betondruckfestigkeit	1	1
42,60	Glaseigenschaften	0	0
41,22	Zusammenstellung der technischen Kennwerte von Nadelhölzern	0	0
40,79	Kurzzeitfestigkeiten von Vollholz	0	0

Σ: 15 17

Suchbegriffe: Beton – Frost

Anzahl Seiten: 10

P: 80,00%

P': 80,00%

R	Titel	R1	R2
89,17	Expositionsklassen hinsichtlich Betonangriff	1	1
82,61	Dauerhaftigkeit von bewehrten Betonkonstruktionen	1	1
79,46	Wirkungsgruppen und Kennzeichnung der Betonzusatzmittel	1	1
78,11	Korrosion der Oberfläche (Betonkorrosion) des Bauteils	1	1
75,53	Dauerhaftigkeit in der Praxis - kombinierte Beanspruchungen	1	1
75,51	Expositionsklassen / XF	1	1
74,57	Betonzusatzmittel / Luftporenbildner (LP)	1	1
74,04	Eigenschaften von Hüttensteinen	0	0
69,49	Ausgangsstoffe des Betons (Zementbeton)	1	1
68,36	Betonzusatzstoffe / inaktive Stoffe	0	0
Σ:		8	8

Suchbegriffe: Beton – Korrosion - Stahl

Anzahl Seiten: 19

P: 63,16%

P': 68,42%

R	Titel	R1	R2
89,17	Einfluss der Temperatur	1	1
83,95	Faserbeton / Stahlfaserarten	0	0
83,70	Definitionen zu Betonstahl	1	1
80,28	Allgemeines zum Bezeichnungssystem für Betonstahl	0	1
79,35	Betonstähle	1	1
78,70	Exkurs Stahlbeton: Depassivierung	1	1
75,66	Allgemeines zum Betonstahl	1	1
75,24	Wirkungsgruppen und Kennzeichnung der Betonzusatzmittel	0	0
73,77	Eigenschaften von Betonstahl	1	1
68,88	Einfluss des pH-Werts	1	1
68,75	Definition Edelstahl	1	1
67,52	Betonzusatzmittel / Einleitung	0	0

66,63	Zusatzstoffe	0	0
66,52	Korrosion der Oberfläche (Betonkorrosion) des Bauteils	1	1
63,72	Betonzusatzstoffe / inaktive Stoffe	0	0
63,65	Zusatzmittel	0	0
62,79	Carbonatisierung / Schadensvermeidung	1	1
58,42	Expositionsklassen hinsichtlich Bewehrungskorrosion	1	1
52,07	Faserbeton / Anwendungsgebiete	1	1
Σ:		12	13

Suchbegriffe: Beton – Kriechen – Schwinden

Anzahl Seiten: 14

P: 57,14%

P': 64,29%

R	Titel	R1	R2
89,17	Trocknungsschwinddehnung	1	1
78,75	Grund- und Trocknungskriechen	1	1
77,83	Kriechen von Beton	1	1
77,74	Relaxation und Kriechen von Kunststoffen	0	0
75,49	Schwinddehnung / Schrumpfdehnung	1	1
71,93	Schwindverformungen	0	0
71,45	Zeitverformungslinie	1	1
69,36	Wirkungsgruppen und Kennzeichnung der Betonzusatzmittel	0	0
65,64	Betonzuschlag / Einführung	1	1
62,38	Langzeitverformungen von Holz - Allgemeines	0	1
55,24	Quell- und Schwindmaße von Holz	0	0
54,79	Lastunabhängige Verformung (Schwinden und Quellen)	1	1
53,99	Schwinden und Quellen von Holz	0	0
53,46	Verformungen durch Kriechen	1	1
Σ:		8	9

Suchbegriffe: Beton – Rohdichte – Zuschlag

Anzahl Seiten: 19

P: 78,95%

P': 84,21%

R	Titel	R1	R2
82,24	Übersicht von verschiedenen Stoffen und deren Dichten	1	1
81,51	Festigkeitsklassen und Rohdichte	1	1

77,36	Schwerbeton / Dichte der Zuschläge	1	1
76,28	Schwerbeton / Zuschlagarten	1	1
73,90	Faserbeton / Zuschlag	1	1
73,90	Sieblinien	1	1
68,81	Ausgangsstoffe des Betons (Zementbeton)	1	1
68,80	Wirkungsgruppen und Kennzeichnung der Betonzusatzmittel	0	0
67,56	Zusammensetzung und Verarbeitung des Betons (Zusatzstoffe, Verdichtung)	0	1
66,33	Einflüsse auf die Konsistenz	1	1
63,32	Schwerbeton / Einführung	1	1
60,61	Einbau des Betons	1	1
60,30	Betonzusatzmittel / Stabilisierer (ST)	1	1
59,51	Schwerbeton / Herstellung	1	1
50,45	Leichtbeton / Definitionen	1	1
46,02	Druckfestigkeitsklassen für Normal- und Schwerbeton	1	1
45,45	Einflüsse auf die Konsistenz	1	1
41,88	Eigenschaften von Mauersteinen aus Normal- und Leichtbeton	0	0
41,73	Betonfestigkeitsklassen	0	0
Σ:		15	16

Suchbegriffe: Beton – Stahl**Anzahl Seiten:** 22**P:** 50,00%**P':** 59,09%

R	Titel	R1	R2
89,17	Allgemeines zum Betonstahl	1	1
84,91	Faserbeton / Stahlfaserarten	1	1
84,52	Definitionen zu Betonstahl	1	1
84,08	Allgemeines zum Bezeichnungssystem für Betonstahl	1	1
79,80	Wirkungsgruppen und Kennzeichnung der Betonzusatzmittel	0	0
79,67	Betonstähle	1	1
78,53	Herstellverfahren für Betonstahl	1	1
74,66	Konsistenz und Temperatur	0	0
74,11	Unlegierte Edelstähle	1	1
72,99	Betonzusatzstoffe / anorganisch-mineralische Zusatzstoffe	0	0
72,82	Betonzusatzstoffe / Einführung	0	0
70,18	Geschichtliche Entwicklung von Betonstahl	1	1

69,81	Edelstähle	0	1
69,77	Betonzusatzmittel / Einleitung	0	0
69,45	Betonzusatzstoffe / inaktive Stoffe	0	0
69,38	Legierte Edelstähle	0	0
68,64	Baustähle und Werkzeugstähle	1	1
68,00	Unlegierte Qualitätsstähle	1	1
67,96	Form und Funktion - Tragwerke aus Beton (2)	1	1
67,14	Allgemeines zu Edelstahl	0	0
67,00	Qualitätsstähle	0	0
66,21	Frischbetoneigenschaften	0	1
Σ:		11	13

Suchbegriffe: Beton – Verwendung – Zusatzmittel**Anzahl Seiten:** 15**P:** 80,00%**P':** 80,00%

R	Titel	R1	R2
89,17	Wirkungsgruppen und Kennzeichnung der Betonzusatzmittel	1	1
89,17	Betonzusatzmittel / Wirkungsgruppen	1	1
89,17	Betonzusatzmittel / Einpresshilfen (EH)	1	1
89,17	Betonzusatzmittel / Dosierung	1	1
83,28	Betonzusatzmittel / Recyclinghilfen für Waschwasser (RH)	1	1
81,44	Ausgangsstoffe des Betons (Zementbeton)	1	1
79,92	Betonzusatzmittel / Fließmittel (FM)	1	1
78,93	Betonzusatzmittel / Luftporenbildner (LP)	1	1
77,99	Betonzusatzmittel / Verzögerer (VZ)	1	1
77,75	Betonzusatzmittel / Einleitung	1	1
69,36	Einflüsse auf die Konsistenz	1	1
62,75	Betonzusatzstoffe / Wirkungsgruppen	0	0
60,9	Zusatzmittel	1	1
52,71	Betonzusatzstoffe / inaktive Stoffe	0	0
48,64	Betonzusatzstoffe / Einführung	0	0
Σ:		12	12

Suchbegriffe: Beton – Zusatzmittel**Anzahl Seiten:** 14**P:** 78,57%**P':** 85,71%

R	Titel	R1	R2
---	-------	----	----

89,17	Wirkungsgruppen und Kennzeichnung der Betonzusatzmittel	1	1
89,17	Betonzusatzmittel / Wirkungsgruppen	1	1
89,17	Betonzusatzmittel / Luftporenbildner (LP)	1	1
89,17	Betonzusatzmittel / Einleitung	1	1
89,17	Betonzusatzmittel / Dosierung	1	1
87,17	Betonzusatzmittel / Einpresshilfen (EH)	1	1
83,08	Betonzusatzmittel / Fließmittel (FM)	1	1
79,73	Betonzusatzmittel / Betonverflüssiger (BV)	1	1
79,07	Betonzusatzmittel / Recyclinghilfen für Waschwasser (RH)	1	1
78,67	Ausgangsstoffe des Betons (Zementbeton)	1	1
77,52	Betonzusatzmittel / Stabilisierer (ST)	1	1
70,10	Betonzusatzstoffe / Wirkungsgruppen	0	0
65,05	Betonzusatzstoffe / inaktive Stoffe	0	0
57,20	Frischbetoneigenschaften	0	1
Σ:		11	12

Suchbegriffe: Bewehrung – Korrosion – pH

Anzahl Seiten: 19

P: 78,95%

P': 94,74%

R	Titel	R1	R2
89,17	Korrosion der Bewehrung	1	1
85,59	Exkurs Stahlbeton: Depassivierung	1	1
78,64	Expositionsklassen hinsichtlich Bewehrungskorrosion	1	1
74,05	Einfluss des pH-Werts	1	1
72,73	Wasserstoffkorrosion	0	1
72,29	Definition des pH-Wertes	1	1
71,52	Dauerhaftigkeit von bewehrten Betonkonstruktionen	1	1
69,65	Chloridkorrosion	1	1
66,99	Sauerstoffkorrosion	1	1
63,13	Anodische und kathodische Teilreaktion	1	1
58,70	Passivität	1	1
58,42	Exkurs Stahlbeton: pH-Wert der Porenlösung	1	1
53,17	Allgemeines zum Bezeichnungssystem für Betonstahl	0	0
50,96	Wasserstoff- und Sauerstoffelektrode	0	1

42,46	Expositionsklassen / Einteilung	1	1
42,36	Zuschlag / Stahlangreifende Stoffe (Salze)	1	1
41,46	Aufgaben und Arten des Korrosionsschutzes	0	1
39,92	Korrosion der Oberfläche (Betonkorrosion) des Bauteils	1	1
36,21	Kathodische Teilreaktion	1	1
Σ:		15	18

Suchbegriffe: Bruch – Stahlbeton

Anzahl Seiten: 14

P: 42,86%

P': 64,29%

R	Titel	R1	R2
77,60	Brucheinschnürung (Z) für einen kaltverformten Stahl	0	0
75,01	Brucheinschnürung (Z) für einen warmverformten Stahl	0	0
74,21	Brucheinschnürung (Z)	0	0
71,69	Bruchdehnung (A)	1	1
70,61	Bruch infolge Ermüdung	1	1
70,3	Einflüsse auf die Bruchzähigkeit	0	0
69,95	Sprödes Werkstoffversagen	1	1
68,95	Spannungs-Dehnungs-Diagramm eines kaltverformten Stahls	1	1
68,39	Bruchdehnung	1	1
68,37	Makroskopisches Bruchverhalten	0	1
68,17	Bruchdehnung (A) beim kaltverformten Stahl	0	1
68,02	Spannungs-Dehnungs-Diagramm eines warmverformten Stahls	1	1
67,42	Bruchdehnung (A) beim warmverformten Stahl	0	1
58,39	Gleichmaßdehnung (Ag)	0	0
Σ:		6	9

Suchbegriffe: Dauer – Erhärten - Frischbeton

Anzahl Seiten: 9

P: 44,44%

P': 100,00%

R	Titel	R1	R2
63,33	Konsistenzbestimmung	0	1
63,33	Erhärten des Betons	1	1
62,86	Konsistenz und Temperatur	0	1

62,01	Ansteifen, Erstarren & Erhärten	1	1
61,15	Einflüsse auf die Konsistenz	1	1
54,13	Konsistenzbereiche	0	1
53,54	Frischbetoneigenschaften	0	1
52,28	Verdichtungsversuch	0	1
49,67	Konsistenz und Fließmittel	1	1
Σ:		4	9

Suchbegriffe: Definition – Dichte

Anzahl Seiten: 14

P: 28,57%

P': 42,86%

R	Titel	R1	R2
89,17	Definition der Rein- und Rohdichte	1	1
79,60	Übersicht von verschiedenen Stoffen und deren Dichten	1	1
79,05	Normalspannungsreihe	0	0
70,79	Neue Möglichkeiten zur Definition des Kilogramms	0	0
69,16	Die technische Festlegung des Amperes	0	0
68,12	Abhängigkeit der Festigkeit von der Rohdichte	0	0
64,56	Das Ampere	0	0
64,53	Festigkeitsklassen und Rohdichte	0	0
64,06	Die Dichte ρ	1	1
63,90	Einführung verschiedener Dichtebegriffe	1	1
60,87	Schwerbeton / Dichte der Zuschläge	0	0
55,49	Ausgewählte Kunststoffe und deren Eigenschaften	0	0
50,00	SI-Basiseinheiten	0	1
50,00	Das Kilogramm	0	1
Σ:		4	6

Suchbegriffe: Druckfestigkeit – Zugfestigkeit

Anzahl Seiten: 16

P: 87,50%

P': 100,00%

R	Titel	R1	R2
78,55	Betondruckfestigkeitsprüfung / Filme	1	1
77,83	Druckfestigkeitsklassen für Normal- und Schwerbeton	1	1
77,83	Druckfestigkeitsklassen für Leichtbeton	1	1
77,02	Spannungs-Dehnungs-Diagramm eines	1	1

warmverformten Stahls

75,63	Faserbeton / Verhalten bei Zug- und Biegebeanspruchung	0	1
72,98	Betondruckfestigkeit / Prüfkörpergestalt und -abmessungen	1	1
72,83	Biegezugfestigkeit	1	1
72,45	Betondruckfestigkeitsprüfung	1	1
72,03	Zentrische Zugfestigkeit	1	1
70,28	Spannungs-Dehnungs-Diagramm eines kaltverformten Stahls	1	1
68,89	Zusammenhang zwischen Zug- und Druckfestigkeit	1	1
67,62	Druckfestigkeit	1	1
65,51	Zugfestigkeit	1	1
63,57	Glaseigenschaften	0	1
63,33	Einflüsse auf die Druckfestigkeit von Beton	1	1
62,58	Betonfestigkeitsklassen	1	1
Σ:		14	16

Suchbegriffe: duktil - Versagen

Anzahl Seiten: 13

P: 84,62%

P': 84,62%

R	Titel	R1	R2
89,17	Duktilen Versagen von Stahlbetonbauteilen mit normalem Bewehrungsgehalt	1	1
89,17	Duktilen Bauteilversagen	1	1
88,72	Robustheit der Konstruktion	1	1
83,30	Schlagartiges Versagen eines Stahlbetonbauteils in der Druckzone	1	1
83,28	Plötzliches Versagen eines Stahlbetonbauteils mit zu geringer Mindestbewehrung	1	1
82,18	Sprödes Werkstoffversagen	1	1
81,63	Sprödes Bauteilversagen	1	1
77,96	Makroskopisches Bruchverhalten	1	1
74,89	Stabilitätsversagen	1	1
71,84	Versagensarten	1	1
69,96	Materialbruch	1	1
61,55	Kippen	0	0
34,70	<u>Abgrenzung zur Ermüdungsmechanik</u>	0	0
Σ:		11	11

Suchbegriffe:	Duroplast – Elastomer - Thermoplast
Anzahl Seiten:	15
P:	80,00%
P':	100,00%

R	Titel	R1	R2
85,75	Einteilung nach dem Herstellungsprinzip	1	1
82,83	Verarbeiten der "polyadditiven" Duroplaste und der Elastomere	1	1
76,87	Einteilung nach der Molekularstruktur	1	1
76,49	Thermoplaste, Duroplaste und Elastomere	1	1
71,33	Grundlagen zum Herstellen von geformten Kunststoffteilen	1	1
70,98	Urformen von duroplastischen Kunststoffen	1	1
68,73	Ausgewählte Kunststoffe und deren Eigenschaften	1	1
63,87	Temperaturabhängigkeit der Kunststoffe	1	1
62,00	Übersicht möglicher Kunststoff-Anwendungen im Bauwesen	0	1
62,00	Übersicht häufiger Kunststoff-Anwendungen im Bauwesen	0	1
53,32	Allgemeine Formgebungsverfahren - Urformen	1	1
52,04	Elastomere	1	1
50,17	Thermoelaste	1	1
50,00	Zeitschne der geschichtlichen Entwicklung von Kunststoffe	1	1
50,00	Kurzzeichen für Kunststoffe	0	1
Σ:		12	15

Suchbegriffe:	Duroplast – Thermoplast
Anzahl Seiten:	14
P:	78,57%
P':	100,00%

R	Titel	R1	R2
88,27	Einteilung nach dem Herstellungsprinzip	1	1
84,99	Einteilung nach der Molekularstruktur	1	1
81,18	Verarbeiten der "polyadditiven" Duroplaste und der Elastomere	1	1
77,79	Thermoplaste, Duroplaste und Elastomere	1	1
77,69	Urformen von duroplastischen Kunststoffen	1	1

77,10	Übersicht möglicher Kunststoff-Anwendungen im Bauwesen	0	1
77,10	Übersicht häufiger Kunststoff-Anwendungen im Bauwesen	0	1
77,03	Grundlagen zum Herstellen von geformten Kunststoffteilen	1	1
66,53	Allgemeine Formgebungsverfahren - Urformen	1	1
61,65	Matrixwerkstoffe	1	1
60,78	Urformen von Thermoplasten (Plastomere)	1	1
53,54	Ausgewählte Kunststoffe und deren Eigenschaften	1	1
50,00	Kurzzeichen für Kunststoffe	0	1
42,10	Zeitschne der geschichtlichen Entwicklung von Kunststoffe	1	1
Σ:		11	14

Suchbegriffe:	Edelstahl – Stahlbau – Verwendung
Anzahl Seiten:	14
P:	50,00%
P':	71,43%

R	Titel	R1	R2
84,02	Legierte Qualitätsstähle	0	0
74,67	Edelstähle	0	1
74,58	Edelstahl im Bauwesen	1	1
72,73	Legierte Edelstähle	1	1
63,49	Unlegierte Edelstähle	1	1
61,77	Baustähle und Werkzeugstähle	1	1
57,37	Allgemeines zu Edelstahl	1	1
46,17	Stahlbaustähle - Feinkornbaustähle	1	1
46,07	Stahlbaustähle - Allgemeines	0	1
43,57	Einsatzgebiete von Aluminium	0	0
39,09	Unlegierte Qualitätsstähle	0	0
35,42	Geschichte der Produktion von Eisen, Gusseisen und Stahl	0	1
30,89	Allgemeines	1	1
30,52	Grundstähle	0	0
Σ:		7	10

Suchbegriffe:	elastisch – plastisch – Verhalten
Anzahl Seiten:	20
P:	90,00%
P':	95,00%

R	Titel	R1	R2
---	-------	----	----

77,11	Das Elastische Verhalten	1	1
69,57	Elastische Verformung	1	1
67,36	Werkstoffverhalten bei Dehnungsumkehr und Einteilung der Werkstoffgesetze	1	1
66,91	Temperaturabhängigkeit der Kunststoffe	1	1
66,66	Das ideale Verformungsverhalten	1	1
64,19	Lösung für den vollplastischen Zustand	1	1
63,90	Viskoelastizität	0	1
63,70	Lösung für linearelastisches Werkstoffgesetz	1	1
63,61	Elastische Verformung beim warmverformten Stahl	1	1
63,09	Linearelastische und nichtlineare Grenzzustände	1	1
61,91	Elastische Verformung beim kaltverformten Stahl	1	1
61,82	Gesamtverformung von Kunststoffen	1	1
60,00	Spannungs-Dehnungs-Diagramme	1	1
59,56	WiBA-Net >> Werkstoffe im Bauwesen und Architektur	1	1
59,01	Werkstoffgesetz	1	1
56,24	Verschiebung der Neutralen Faser bei nichtlinearem Werkstoffgesetz	1	1
54,53	Das Plastische Verhalten	1	1
54,12	Einflussgrößen	1	1
49,14	Lösungsgleichungen	0	0
43,37	Spannungs-Dehnungs-Verhalten von Stahl ohne ausgeprägter Steckgrenze	1	1
Σ:		18	19

Suchbegriffe: Elastomer – Thermoplast

Anzahl Seiten: 17

P: 76,47%

P': 94,12%

R	Titel	R1	R2
89,17	Einteilung nach dem Herstellungsprinzip	1	1
86,56	Einteilung nach der Molekularstruktur	1	1
84,38	Thermoelaste	1	1
81,21	Thermoplaste, Duroplaste und Elastomere	1	1
77,91	Grundlagen zum Herstellen von geformten Kunststoffteilen	1	1
77,43	Temperaturabhängigkeit der Kunststoffe	1	1

75,51	Ausgewählte Kunststoffe und deren Eigenschaften	1	1
75,11	Verarbeiten der "polyadditiven" Duroplaste und der Elastomere	1	1
74,68	Elastomere	1	1
65,32	Allgemeine Formgebungsverfahren - Urformen	1	1
64,36	Systematik der Klebstoffe in Anlehnung an DIN 16 920	0	0
63,00	Elastomer- und Plastomerbitumen	1	1
62,82	Urformen von Thermoplasten (Plastomere)	1	1
50,00	Kurzzeichen für Kunststoffe	0	1
50,00	Übersicht möglicher Kunststoff-Anwendungen im Bauwesen	0	1
50,00	Übersicht häufiger Kunststoff-Anwendungen im Bauwesen	0	1
38,95	Zeitschne der geschichtlichen Entwicklung von Kunststoffe	1	1
Σ:		13	16

Suchbegriffe: Erhärten – Frischbeton

Anzahl Seiten: 12

P: 58,33%

P': 100,00%

R	Titel	R1	R2
84,14	Erhärten des Betons	1	1
81,62	Konsistenzbestimmung	0	1
77,36	Konsistenz und Temperatur	1	1
72,13	Frischbetoneigenschaften	0	1
71,41	Einflüsse auf die Konsistenz	0	1
66,52	Konsistenz (des Frischbetons)	1	1
64,85	Betonzusatzmittel / Verzögerer (VZ)	1	1
63,10	Berechnung der Frischbetontemperatur	0	1
60,59	Konsistenzbereiche	1	1
58,90	Konsistenz und Fließmittel	1	1
57,85	Ansteifen, Erstarren & Erhärten	1	1
56,07	Einflüsse auf die Konsistenz	0	1
Σ:		7	12

Suchbegriffe: Erzeugung – Hochofen - Stahl

Anzahl Seiten: 15

P: 80,00%

P': 80,00%

R	Titel	R1	R2
---	-------	----	----

89,17	Direktreduktion	1	1
87,56	Brennstoffe	1	1
75,06	Hochofenprozess	1	1
71,90	Verfestigungsbereich beim kaltverformten Stahl	0	0
71,73	Winderhitzer (Cowper)	1	1
70,00	Geschichte der Produktion von Eisen, Gusseisen und Stahl	1	1
69,81	Stahlerzeugung heute	1	1
69,35	Aufbau des Hochofens	1	1
69,31	Verfestigungsbereich beim warmverformten Stahl	0	0
67,54	Nachentschwefelung	1	1
63,33	Massebilanz	1	1
63,33	Gichtgas	1	1
62,32	Hauptarten des Zements	0	0
59,04	Boudouard-Gleichgewicht	1	1
51,80	Zuschlag	1	1
Σ:		12	12

Suchbegriffe: Erzeugung - Stahl

Anzahl Seiten: 17

P: 23,53%

P': 47,06%

R	Titel	R1	R2
79,58	Geschichte der Produktion von Eisen, Gusseisen und Stahl	1	1
77,92	Edelstähle	0	0
76,93	Stahlerzeugung heute	1	1
76,90	Verfestigungsbereich beim kaltverformten Stahl	0	0
74,07	Definition Edelstahl	0	0
72,25	Verfestigungsbereich beim warmverformten Stahl	0	0
70,83	Allgemeines zu Edelstahl	0	1
70,34	Legierte Edelstähle	0	0
70,19	Spannstähle	0	1
69,74	Baustähle und Werkzeugstähle	1	1
66,55	Betonstähle	1	1
66,32	Allgemeines zum Betonstahl	0	1
65,91	Unlegierte Qualitätsstähle	0	0
65,90	Allgemeines zum Bezeichnungssystem für Betonstahl	0	0
64,75	Allgemeines zu Spannstählen	0	1

57,99	Definitionen zu Betonstahl	0	0
57,80	Verfestigungsbereich	0	0
Σ:		4	8

Suchbegriffe: Festigkeit – Legierung – Stahl

Anzahl Seiten: 22

P: 54,55%

P': 77,27%

R	Titel	R1	R2
84,06	Legierte Edelstähle	1	1
83,76	Unlegierte Stähle	0	1
82,26	Werkstoffbezeichnungen des Gusseisens mit Lamellengraphit	0	0
81,76	Werkstoffbezeichnungen des Gusseisens mit Kugelgraphit	0	0
81,67	Definition Edelstahl	1	1
81,20	Hochlegierte Stähle	1	1
80,53	Unlegierte Stähle mit Mn > 1% und legierte Stähle mit Legierungsgehalten < 5% (niedrig legierte Stähle)	1	1
80,43	Legierte Qualitätsstähle	1	1
79,88	Druckfestigkeitsklassen für Normal- und Schwerbeton	0	0
78,09	Niedriglegierte Stähle	1	1
77,15	Druckfestigkeitsklassen für Leichtbeton	0	0
73,93	Spannungs-Dehnungs-Diagramm eines kaltverformten Stahls	1	1
72,19	Unlegierte Edelstähle	0	1
71,97	Andere legierte Stähle	1	1
71,07	Qualitätsstähle	1	1
70,11	Unlegierte Qualitätsstähle	1	1
68,52	Spannungs-Dehnungs-Diagramm eines warmverformten Stahls	1	1
66,48	Qualitätsstähle in anderen Bereichen - Einsatzstähle	0	1
64,41	Dauerfestigkeit und Grenzlasterzahl	0	0
64,25	Allgemeines zu Edelstahl	1	1
61,74	Edelstähle	0	1
60,49	Grundstähle	0	1
Σ:		12	17

Suchbegriffe: Feuchte – Holz

Anzahl Seiten: 17

P: 82,35%

P': 88,24%

R	Titel	R1	R2	Holzschutzes - Begriffsklärung			
89,17	Gleichgewichtsfeuchte von Holz	1	1	75,42	Sorptionsisotherme und Mechanismen der Feuchteaufnahme	1	1
79,97	Baulicher Holzschutz	1	1	74,72	Konstruktiver Holzschutz - Schutz gegen Feuchteleitung	1	1
79,65	Konstruktiver Holzschutz - feuchtegefährdete Holzbauteile	1	1	72,13	Gefährdungsklassen von Holzbauteilen nach DIN 68 800	0	1
79,39	Gefährdungsklasse 2	1	1	70,09	Schwindverformungen	1	1
78,94	Gefährdungsklassen von Holzbauteilen nach DIN 68 800	1	1	56,12	Trocknen von Holz	1	1
78,61	Einfluss der Holzfeuchte auf das Verformungsverhalten	1	1	53,85	Relevante Eigenschaften der für Bauzwecke eingesetzten Holzarten	1	1
77,82	Gefährdungsklasse 1	1	1	45,65	Zeitschiene der Entwicklung des Holzbaus	0	0
77,29	Endkriechzahl für Normalbeton und feuchte Umgebungsbedingungen	0	0	Σ: 12 14			
75,73	Holzfeuchte	1	1	Suchbegriffe: flächenzentriert – kubisch Anzahl Seiten: 7 P: 85,71% P': 85,71%			
75,12	Holzwerkstoffklassen	1	1				
75,11	Grundprinzipien des Konstruktiven Holzschutzes - Begriffsklärung	1	1				
74,07	Einfluss der Temperatur auf das Verformungsmodul von Holz	0	1				
72,80	Konstruktiver Holzschutz - Schutz gegen Feuchteleitung	1	1	R	Titel	R1	R2
71,32	Quell- und Schwindmaße von Holz	1	1	89,17	Kubisch raumzentriert (krz)	1	1
70,41	Zeitschiene der Entwicklung des Holzbaus	0	0	89,17	Kubisch flächenzentriert (kfz)	1	1
69,17	Maßnahmen des Holzschutzes	1	1	89,17	Gleitsysteme im kubisch flächenzentrierten System	1	1
66,09	Gefährdungsklasse 0	1	1	89,17	Bravais-Gitter	0	0
Σ: 14 15				80,50	Gleitsysteme im kubisch raumzentrierten System	1	1
Suchbegriffe: Feuchte – Holz – Quellen Anzahl Seiten: 17 P: 70,59% P': 82,35%				50,00	hdp-Struktur	1	1
				44,56	Gitterkonstanten	1	1
				Σ: 6 6			
				Suchbegriffe: Flussmittel – Löten Anzahl Seiten: 14 P: 92,86% P': 92,86%			
R	Titel	R1	R2	R	Titel	R1	R2
89,17	Schwinden und Quellen von Holz	1	1	89,17	Hartlöten	1	1
89,17	Quell- und Schwindmaße von Holz	1	1	84,78	Hochtemperaturlöten	1	1
89,17	Gleichgewichtsfeuchte von Holz	1	1	81,81	Schutzgase	1	1
89,17	Einfluss der Holzfeuchte auf das Verformungsverhalten	1	1	79,58	Auftragslöten	1	1
80,96	Gefährdungsklasse 2	0	0	79,30	Löten - Rollenlöten	1	1
79,93	Gefährdungsklasse 1	0	0	78,06	Weichlöten	1	1
79,50	Charakteristische Eigenschaften - Schwinden und Quellen	1	1	78,06	Flussmittel	1	1
78,79	Konstruktiver Holzschutz - feuchtegefährdete Holzbauteile	1	1	71,32	Allgemeines	1	1
77,89	Holzfeuchte	1	1	70,54	Lotbadtauchlöten	1	1
76,41	Grundprinzipien des Konstruktiven	0	1				

70,00	Wellen-/ Schwallöten	1	1
68,57	Kolbenlöten	1	1
67,76	Widerstandslöten	1	1
67,62	Induktionslöten	1	1
31,29	<u>Aluminiumprodukte im Bauwesen</u>	0	0
Σ:		13	13

Suchbegriffe: Frost – Luftpore

Anzahl Seiten: 7

P: 71,43%

P': 85,71%

R	Titel	R1	R2
80,3	Expositionsklassen hinsichtlich Beton- angriff	1	1
79,74	Betonzusatzmittel / Luftporenbildner (LP)	1	1
78,41	Künstlich eingeführte Luftporen	1	1
77,75	Betonzusatzmittel / Wirkungsgruppen	1	1
75,74	Betonzusatzmittel / Dichtungsmittel (DM)	0	1
64,80	Eigenschaften von Hüttensteinen	0	0
45,94	Korrosion der Oberfläche (Betonkorro- sion) des Bauteils	1	1
Σ:		5	6

Suchbegriffe: Gleichgewichtsfeuchte – Holz – Quellen

Anzahl Seiten: 10

P: 90,00%

P': 100,00%

R	Titel	R1	R2
89,17	Schwinden und Quellen von Holz	1	1
89,17	Quell- und Schwindmaße von Holz	1	1
89,17	Gleichgewichtsfeuchte von Holz	1	1
82,25	Sorptionsisotherme und Mechanismen der Feuchteaufnahme	1	1
72,58	Einfluß der Holzfeuchte auf das Verfor- mungsverhalten	1	1
71,46	Charakteristische Eigenschaften - Schwinden und Quellen	1	1
61,53	Trocknen von Holz	1	1
56,57	Schwindverformungen	1	1
52,22	Relevante Eigenschaften der für Bau- zwecke eingesetzten Holzarten	1	1
38,95	Einfluß der Temperatur auf das Verfor- mungsmodul von Holz	0	1

Σ: 9 10

Suchbegriffe: Herstellung – Hochofen – Roheisen

Anzahl Seiten: 13

P: 84,62%

P': 84,62%

R	Titel	R1	R2
89,17	Massebilanz	1	1
83,92	Direktreduktion	1	1
77,14	Brennstoffe	1	1
70,84	Hochofenprozess	1	1
70,00	Geschichte der Produktion von Eisen, Gusseisen und Stahl	1	1
62,16	Schlacke	1	1
59,80	Zusammenfassung: Stahlherstellung	1	1
56,54	Betonzusatzstoffe / Latent-hydraulische Stoffe	0	0
54,74	Nachentschwefelung	1	1
54,34	Erze	1	1
49,68	Zuschlag	1	1
49,56	Gichtgas	1	1
45,53	Verfahren der Zementherstellung	0	0
Σ:		11	11

Suchbegriffe: Herstellung – Hochofen – Roheisen –
Stahl

Anzahl Seiten: 15

P: 80,00%

P': 80,00%

R	Titel	R1	R2
89,17	Massebilanz	1	1
83,77	Direktreduktion	1	1
81,22	Schliffbildherstellung (ungeätzt)	0	0
72,12	Brennstoffe	1	1
69,51	Zusammenfassung: Stahlherstellung	1	1
64,26	Hochofenprozess	1	1
64,14	Nachentschwefelung	1	1
58,00	Gichtgas	1	1
57,83	Geschichte der Produktion von Eisen, Gusseisen und Stahl	1	1
56,87	Schlacke	1	1
55,05	Legierte Edelstähle	0	0
54,36	Betonzusatzstoffe / Latent-hydraulische Stoffe	0	0

52,33	Winderhitzer (Cowper)	1	1
46,98	Zuschlag	1	1
39,69	Erze	1	1
Σ:		12	12

Suchbegriffe: Herstellung – Hochofen – Stahl

Anzahl Seiten: 15

P: 80,00%

P': 80,00%

R	Titel	R1	R2
89,17	Schliffbildherstellung (ungeätzt)	0	0
83,09	Massebilanz	1	1
76,83	Zusammenfassung: Stahlherstellung	1	1
70,27	Hochofenprozess	1	1
70,00	Geschichte der Produktion von Eisen, Gusseisen und Stahl	1	1
64,65	Winderhitzer (Cowper)	1	1
60,02	Nachentschwefelung	1	1
58,87	Legierte Edelstähle	0	0
58,51	Brennstoffe	1	1
58,41	Gichtgas	1	1
55,69	Edelstähle	0	0
52,84	Direktreduktion	1	1
51,71	Schlacke	1	1
51,26	Zuschlag	1	1
41,57	Erze	1	1
Σ:		12	12

Suchbegriffe: Herstellung – Stahl

Anzahl Seiten: 15

P: 60,00%

P': 66,67%

R	Titel	R1	R2
79,67	Zusammenfassung: Stahlherstellung	1	1
76,44	Schliffbildherstellung (ungeätzt)	0	0
74,21	Herstellverfahren für Betonstahl	1	1
72,60	Definitionen zu Betonstahl	1	1
72,22	Unlegierte Edelstähle	1	1
71,33	Allgemeines zum Bezeichnungssystem für Betonstahl	0	0
71,29	Edelstähle	1	1
70,96	Legierte Edelstähle	1	1
66,09	Geschichtliche Entwicklung von Beton-	1	1

stahl

65,71	Allgemeines zum Betonstahl	0	0
64,74	Allgemeines zu Edelstahl	1	1
63,87	Unlegierte Qualitätsstähle	0	0
63,85	Betonstähle	0	1
63,58	Baustähle und Werkzeugstähle	1	1
62,99	Stahlbaustähle - Allgemeines	0	0
Σ:		9	10

Suchbegriffe: Herstellung – Zement

Anzahl Seiten: 14

P: 64,29%

P': 71,43%

R	Titel	R1	R2
89,17	Verfahren der Zementherstellung	1	1
86,2	Das Verfahren der Zementherstellung	1	1
79,58	Erhärten des Betons	0	0
77,05	Zement	1	1
75,6	Hauptbestandteile des Zement	1	1
74,37	Hauptbestandteile	1	1
73,22	Hauptarten des Zements	0	1
72,01	Ausgangsstoffe des Betons (Zementbeton)	0	0
70,92	Portlandzementklinker (K)	1	1
67,00	Brennen des Zements	1	1
65,76	Schliffbildherstellung (ungeätzt)	0	0
57,97	Begriffsbestimmung - Zement	1	1
57,47	Zusammenfassung: Stahlherstellung	0	0
57,08	Ausgangsstoffe	1	1
Σ:		9	10

Suchbegriffe: Hochofen – Stahlerzeugung

Anzahl Seiten: 9

P: 77,78%

P': 88,89%

R	Titel	R1	R2
85,39	Stahlerzeugung heute	1	1
84,38	Massebilanz	1	1
83,21	Das Puddelverfahren seit 1783	0	1
73,21	Geschichte der Produktion von Eisen, Gusseisen und Stahl	1	1
73,03	Hochofenprozess	1	1
69,64	Winderhitzer (Cowper)	1	1

69,38	Gichtgas	1	1
68,15	Aufbau des Hochofens	1	1
66,43	Vergleich alte und neue Norm	0	0
Σ:		8	9

Suchbegriffe: Holz – Verbindung

Anzahl Seiten: 19

P: 63,16%

P': 68,42%

R	Titel	R1	R2
89,17	Einführung - Handwerkliche Holzverbindungen	1	1
88,82	Moderne Holzverbindungen mit Stahlelementen	1	1
87,72	(3) Hölzerne Verbindungsmittel	1	1
77,07	Holzverbindungen	1	1
74,68	Zangenanschlüsse - Einsatzorte	1	1
73,22	Zeitschne der Entwicklung des Holzbau	1	1
72,84	Verbindungsmittel - Moderne Verbindungstechnik mit Stahlelementen	1	1
69,96	Allgemeines zu Verbindungsarten	0	1
69,00	Allgemeines zu Verbindungsmittel (VBM)	1	1
68,93	Trocknen von Holz	0	0
68,65	Baulicher Holzschutz	0	0
68,51	Pressdübel - Einlassdübel & Einpressdübel	1	1
68,44	Grundprinzipien des Konstruktiven Holzschutzes - Begriffsklärung	0	0
68,24	Anwendungsbeispiel - Hallenbau mit Brettschichtholz (BSH)	1	1
67,93	Nadelhölzer	0	0
67,58	Gefährdungsklassen von Holzbauteilen nach DIN 68 800	0	0
67,40	Der Baustoff Holz - seine charakteristischen Eigenschaften	1	1
67,36	Konstruktiver Holzschutz - feuchtegefährdete Holzbauteile	0	0
67,11	Konstruktiver Holzschutz - Schutz gegen Feuchteleitung	1	1
Σ:		12	13

Suchbegriffe: Hydratation – Zement

Anzahl Seiten: 11

P: 90,91%

P': 100,00%			
R	Titel	R1	R2
84,38	Erhärten des Betons	1	1
77,86	Schnelle Hydratationswärmeentwicklung	1	1
76,89	Hydratationsstufe III	1	1
74,29	Hauptbestandteile des Zement	1	1
73,19	Hydratationswärme	1	1
71,60	Hauptarten des Zements	0	1
70,97	Hydratationsbeeinflussende Stoffe	1	1
68,76	Hydratation	1	1
68,04	Hauptbestandteile	1	1
66,98	Zement	1	1
64,59	Ansteifen, Erstarren & Erhärten	1	1
Σ:		10	11

Suchbegriffe: Korrosion – Metall

Anzahl Seiten: 17

P: 82,35%

P': 82,35%

R	Titel	R1	R2
86,26	Kontaktkorrosion (Bimetallkorrosion)	1	1
75,82	Arten und Erscheinungen der Metallkorrosion	1	1
74,66	Übersicht wichtiger Nichteisenmetalle (1)	0	0
74,31	Chloridkorrosion	1	1
73,13	Kathodische Fremdmetalüberzüge	1	1
72,24	Korrosion der Bewehrung	1	1
71,72	Schmelztauchüberzüge (Feuerverzinken)	1	1
71,71	Anodische Fremdmetalüberzüge	1	1
70,36	Sauerstoffkorrosion	1	1
69,35	Lochkorrosion	1	1
69,29	Potentialunterschied	1	1
69,22	Interkristalline Spannungsrisskorrosion	1	1
69,18	Einteilungsprinzipien für Nichteisenmetalle	0	0
68,01	Aufgaben und Arten des Korrosionsschutzes	1	1
67,37	Formbarkeit von NE-Metallen	0	0
66,94	Spannungsrisskorrosion	1	1
54,25	Anodische und kathodische Teilreaktion	1	1

Σ: 14 14			
Suchbegriffe: Korrosion – pH – Stahl Anzahl Seiten: 12 P: 75,00% P': 100,00%			
R	Titel	R1	R2
79,52	Wasserstoffkorrosion	0	1
79,2	Einfluss der Temperatur	0	1
77,18	Exkurs Stahlbeton: pH-Wert der Porenlösung	1	1
76,8	Exkurs Stahlbeton: Depassivierung	1	1
75,01	Einfluss des pH-Werts	1	1
68,49	Sauerstoffkorrosion	1	1
65,70	Wasserstoffentstehung bei der Korrosion	0	1
62,32	Anodische und kathodische Teilreaktion	1	1
54,1	Chloridkorrosion	1	1
51,5	Korrosion der Bewehrung	1	1
49,16	Expositionsklassen hinsichtlich Bewehrungskorrosion	1	1
39,75	Kathodische Teilreaktion	1	1
Σ: 9		12	

Suchbegriffe: Kunststoff – Herstellung Anzahl Seiten: 15 P: 66,67% P': 80,00%			
R	Titel	R1	R2
78,09	Kunststoffe	1	1
76,98	Einteilung nach dem Herstellungsprinzip	1	1
76,62	Zeitschiene der geschichtlichen Entwicklung von Kunststoffe	1	1
70,85	Kunststoffproduktion	1	1
70,03	Vom schwarzen Bakelit zu transparenten High-Tech-Kunststoffen (2)	1	1
69,75	Zusammenfassung: Stahlherstellung	0	0
69,52	Schliffbildherstellung (ungeätzt)	0	0
68,74	Übersicht möglicher Kunststoff-Anwendungen im Bauwesen	0	1
68,66	Einführung - Was sind Kunststoffe?	1	1
68,58	Spannungs-Dehnungs-Diagramme	1	1
68,56	Übersicht häufiger Kunststoff-Anwen-	0	1

dungen im Bauwesen			
66,27	Kunststoffe - Einführung	1	1
65,68	Kunststoffüberzüge	0	0
63,57	Vom schwarzen Bakelit zu transparenten High-Tech-Kunststoffen (3)	1	1
56,39	Kurzzeichen für Kunststoffe	1	1
Σ: 10		12	

Suchbegriffe: Kunststoff – Polymerisation Anzahl Seiten: 14 P: 71,43% P': 100,00%			
R	Titel	R1	R2
89,17	Einteilung nach dem Herstellungsprinzip	1	1
83,03	Kunststoffe	1	1
80,72	Vergleich: Polymerisation-Polykondensation-Polyaddition	1	1
77,02	Polykondensation	0	1
76,98	Reaktion vom Monomeren zum Polymeren	1	1
75,05	Zeitschiene der geschichtlichen Entwicklung von Kunststoffe	0	1
74,36	Übersicht möglicher Kunststoff-Anwendungen im Bauwesen	1	1
74,36	Übersicht häufiger Kunststoff-Anwendungen im Bauwesen	1	1
71,78	Variation der Netzkettlänge bei vernetzten Kunststoffen	1	1
71,27	Polymerisation - radikalisch	1	1
71,09	Einführung - Was sind Kunststoffe?	1	1
70,23	Kunststoffe - Einführung	1	1
69,57	Kurzzeichen für Kunststoffe	0	1
59,5	Ausgewählte Kunststoffe und deren Eigenschaften	0	1
Σ: 10		14	

Suchbegriffe: Legierung – Stahl Anzahl Seiten: 20 P: 60,00% P': 85,00%			
R	Titel	R1	R2
86,43	Unlegierte Stähle	1	1
85,08	Legierte Edelstähle	0	1
84,38	Grundstähle	0	0

80,81	Unlegierte Edelstähle	0	1
79,95	Unlegierte Stähle mit Mn > 1% und legierte Stähle mit Legierungsgehalten < 5% (niedrig legierte Stähle)	1	1
78,99	Allgemeines zu Edelstahl	1	1
78,56	Hochlegierte Stähle	1	1
78,15	Legierungen	1	1
76,76	Andere legierte Stähle	1	1
76,29	Definition Edelstahl	1	1
73,57	Qualitätsstähle	0	1
72,91	Legierte Qualitätsstähle	1	1
71,38	Edelstähle	1	1
70,47	Zink-Legierungen	1	1
69,71	Unlegierte Qualitätsstähle	0	1
68,91	Stahlbaustähle - Allgemeines	0	1
68,28	Qualitätsstähle in anderen Bereichen - Einsatzstähle	0	0
68,14	Nichtrostende Stähle	1	1
67,97	Aluminium-Legierungen	1	1
65,61	Definitionen zu Betonstahl	0	0
Σ:		12	17

Suchbegriffe: Legierungselement – Mangan
Anzahl Seiten: 14
P: 78,57%
P': 92,86%

R	Titel	R1	R2
79,58	Unlegierte Stähle	1	1
79,58	Andere legierte Stähle	1	1
79,42	Stahlbegleiter	1	1
78,72	Legieren/Legierungselemente	1	1
75,88	Einfluss der Legierungselemente	1	1
74,6	Sekundärmetallurgie	1	1
70,66	Hochlegierte Stähle	1	1
67,94	Bezeichnung mit Kurznamen nach chemischer Zusammensetzung	1	1
67,07	Stahlherstellung	1	1
63,59	Unlegierte Stähle mit Mn < 1%	0	1
62,08	Übersicht des Einflusses der Legierungselemente auf die Schweißbarkeit	1	1
60,58	Frischen	1	1
55,66	Definition Edelstahl	0	0
39,49	Unlegierte Stähle mit Mn > 1% und legierte Stähle mit Legierungsgehalten <	0	1

5% (niedrig legierte Stähle)

Σ: 11 13

Suchbegriffe: Polyaddition – Polykondensation
Anzahl Seiten: 10
P: 90,00%
P': 90,00%

R	Titel	R1	R2
89,17	Vergleich: Polymerisation-Polykondensation-Polyaddition	1	1
89,17	Einteilung nach dem Herstellungsprinzip	1	1
80,10	Polykondensation	1	1
80,05	Polyaddition	1	1
77,79	Reaktion vom Monomeren zum Polymeren	1	1
70,86	Einkomponentenklebstoffe	1	1
69,78	Werkstoffe zur Verklebung mit Zweikomponentenklebern	0	0
67,33	Einführung - Was sind Kunststoffe?	1	1
64,82	Polykondensation am Beispiel von Phenolharz	1	1
62,61	Verarbeiten der Duroplaste-Polykondensationsprodukte	1	1
Σ:		9	9

Suchbegriffe: Polyaddition – Polykondensation – Polymerisation
Anzahl Seiten: 12
P: 83,33%
P': 91,67%

R	Titel	R1	R2
89,17	Vergleich: Polymerisation-Polykondensation-Polyaddition	1	1
89,17	Einteilung nach dem Herstellungsprinzip	1	1
84,66	Polykondensation	1	1
82,05	Reaktion vom Monomeren zum Polymeren	1	1
81,38	Polyaddition	1	1
78,57	Polymerisation - ionisch	1	1
61,71	Verarbeiten der Duroplaste-Polykondensationsprodukte	1	1
57,48	Abbinden mit chemischer Reaktion - Reaktionskleber und ihre Umgebungsbedingungen	0	0
56,35	Polymerisation - radikalisch	1	1

50,00	Kurzzeichen für Kunststoffe	0	1
50,00	Zeitschiene der geschichtlichen Entwicklung von Kunststoffe	1	1
40,13	Kunststoffe	1	1
Σ:		10	11

Suchbegriffe: Polyaddition - Polymerisation

Anzahl Seiten: 12

P: 91,67%

P': 100,00%

R	Titel	R1	R2
89,17	Vergleich: Polymerisation-Polykondensation-Polyaddition	1	1
89,17	Einteilung nach dem Herstellungsprinzip	1	1
81,05	Polyaddition	1	1
80,97	Polykondensation	1	1
80,70	Reaktion vom Monomeren zum Polymeren	1	1
74,64	Polymerisation - ionisch	1	1
72,23	Polymerisation - radikalisch	1	1
65,52	Einführung - Was sind Kunststoffe?	1	1
61,30	Abbinden mit chemischer Reaktion - Reaktionskleber und ihre Umgebungsbedingungen	1	1
53,46	Verarbeiten der Duroplaste-Polykondensationsprodukte	0	1
51,59	Abbindevorgänge beim Kleben	1	1
34,85	Kunststoffe	1	1
Σ:		11	12

Suchbegriffe: Polykondensation – Polymerisation

Anzahl Seiten: 11

P: 90,91%

P': 90,91%

R	Titel	R1	R2
89,17	Vergleich: Polymerisation-Polykondensation-Polyaddition	1	1
89,17	Einteilung nach dem Herstellungsprinzip	1	1
81,08	Reaktion vom Monomeren zum Polymeren	1	1
80,4	Polymerisation - ionisch	1	1
80,39	Polyaddition	1	1
79,63	Polykondensation	1	1

67,48	Werkstoffe zur Verklebung mit Zweikomponentenklebern	0	0
65,22	Polymerisation - radikalisch	1	1
63,51	Verarbeiten der Duroplaste-Polykondensationsprodukte	1	1
59,73	Einführung - Was sind Kunststoffe?	1	1
40,56	Kunststoffe	1	1
Σ:		10	10

Suchbegriffe: Schweißen – Verfahren

Anzahl Seiten: 18

P: 77,78%

P': 88,89%

R	Titel	R1	R2
89,17	Verbindungsschweißen	1	1
76,65	Schweißfehler - Fehler im Schweißnahtübergang	1	1
73,71	Manuelles Schweißen	1	1
73,64	Pressschweißen	1	1
71,69	Schweißfehler - Schlackeneinschlüsse	1	1
71,34	Schweißnahtprüfung - Allgemeines	0	1
71,33	Schweiße Elektroden - Allgemeines	1	1
71,17	Schmelzschweißen	1	1
70,18	Abbrennstumpfschweißen/Gaspressschweißen	1	1
69,81	Schweißbeignung - Kohlenstoffäquivalent CEv	1	1
69,79	Schweißfehler - Wolframeinschlüsse	1	1
69,31	Holzschutzmittel-Nichtdruckverfahren	0	0
67,95	Schweißfehler - Einbrandkerben	0	1
67,95	Gasschweißen/Autogenschweißen	1	1
67,17	Verfahrensbedingte Fehler	1	1
66,99	Schweißbeignung	1	1
65,73	Schweißen - Allgemeines	1	1
49,57	<u>Das Windfrischverfahren</u>	0	0
Σ:		14	16

Anhang K: Regressionsrechnung zu den untersuchten Lehrpfaden

In Tabelle K.1 finden sich die für die Regressionsrechnung erforderlichen Ausgangswerte. Wegen der besseren Übersichtlichkeit sind die Werte für die Ober- und Untergrenze von Prognose- und Konfidenzintervall in der separaten Tabelle K.2 dargestellt. Die Spalten (1) und (2) sind in beiden Fällen identisch.

Der Berechnung wird die Transformation $w = \ln(VS)$ zugrunde gelegt. Der Stichprobenumfang beträgt $n = 62$. Für die Spaltensummen der Spalten (3) bis (7) und die Mittelwerte der Spalten (3) und (4), die für die weitere Berechnung benötigt werden, gilt:

$$\begin{aligned}\Sigma w_i &= -147,37 & w_m &= -2,377 \\ \Sigma y_i &= 4.410,77 & y_m &= 71,141 \\ \Sigma w_i^2 &= 625,30 \\ \Sigma y_i^2 &= 331.092,81 \\ \Sigma w_i \cdot y_i &= -9.055,60\end{aligned}$$

Hieraus kann nach Formel (7.2) der Korrelationskoeffizient r berechnet werden, es gilt:

$$r = \frac{62 \cdot -9055,60 - (-147,37 \cdot 4410,77)}{\sqrt{[62 \cdot 625,30 - (-147,37)^2] \cdot [62 \cdot 331092,81 - (4410,77)^2]}} = 0,655 \quad (\text{K.1})$$

Für den empirischen Regressionskoeffizienten b gilt nach Formel (7.3):

$$b = \frac{62 \cdot -9055,60 - (-147,37 \cdot 4410,77)}{62 \cdot 625,30 - (-147,37)^2} = 5,20 \quad (\text{K.2})$$

Für die y-Achsen-Verschiebung a der Regressionsgeraden gilt nach Formel (7.4):

$$a = \frac{4410,77 \cdot 625,30 - (-147,37 \cdot -9055,60)}{62 \cdot 625,30 - (-147,37)^2} = 83,49 \quad (\text{K.3})$$

Mit dem Regressionskoeffizienten b lassen sich die Schätzwerte y_{i0} an den Stellen w_i berechnen. Aus diesen werden dann die Quadrate der Abweichungen der tatsächlichen von den geschätzten Werten $(y_i - y_{i0})^2$ bestimmt. Für deren Summe gilt:

$$\sum (y_i - y_{i0})^2 = 9.880,86 \quad (\text{K.4})$$

Für als Schätzwert s für die Standardabweichung der Grundgesamtheit gilt nach Formel (7.12) mit $P_i = y_i$ und $P_{i0} = y_{i0}$:

$$s = \sqrt{\frac{1}{60-2} \cdot 9880,86} = 12,83 \quad (\text{K.5})$$

Tabelle K.1: Zusammenstellung der für die Regressionsrechnung für den Zusammenhang zwischen VS und P benötigten Ausgangswerte

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Nr.	VS	$w_i = \ln(\text{VS})$	y_i	w_i^2	y_i^2	$w_i \cdot y_i$	$y_{i0} = y_m + b(w_i - w_m)$	$(y_i - y_{i0})^2$	$(w_i - w_m)^2$
1	0,0009	-7,02	15,38	49,25	236,54	-107,94	47,03	1001,63	21,54
2	0,0018	-6,32	28,57	39,94	816,24	-180,55	50,66	487,82	15,54
3	0,0034	-5,68	68,75	32,30	4726,56	-390,71	53,96	218,63	10,93
4	0,0038	-5,58	42,86	31,16	1836,98	-239,25	54,49	135,24	10,27
5	0,0098	-4,62	54,55	21,36	2975,70	-252,14	59,48	24,27	5,04
6	0,0120	-4,44	23,53	19,75	553,66	-104,57	60,40	1359,58	4,27
7	0,0123	-4,40	80,00	19,34	6400,00	-351,78	60,64	374,62	4,08
8	0,0135	-4,30	58,82	18,50	3459,79	-253,01	61,14	5,39	3,70
9	0,0142	-4,26	60,00	18,12	3600,00	-255,42	61,37	1,89	3,53
10	0,0142	-4,26	64,29	18,11	4133,20	-273,57	61,38	8,45	3,53
11	0,0143	-4,25	80,00	18,02	6.400,00	-339,60	61,44	344,63	3,49
12	0,0181	-4,01	63,16	16,10	3.989,19	-253,47	62,64	0,27	2,68
13	0,0182	-4,00	63,16	16,04	3.989,19	-252,95	62,68	0,23	2,65
14	0,0185	-3,99	60,00	15,91	3.600,00	-239,32	62,77	7,66	2,60
15	0,0198	-3,92	75,00	15,38	5.625,00	-294,11	63,12	141,21	2,39
16	0,0224	-3,80	80,00	14,42	6.400,00	-303,79	63,76	263,68	2,02
17	0,0227	-3,78	50,00	14,32	2.500,00	-189,20	63,83	191,31	1,98
18	0,0245	-3,71	71,43	13,75	5.102,24	-264,90	64,22	51,93	1,77
19	0,0260	-3,65	60,00	13,32	3.600,00	-218,94	64,53	20,54	1,62
20	0,0262	-3,64	80,00	13,27	6.400,00	-291,47	64,56	238,35	1,60
21	0,0343	-3,37	66,67	11,38	4.444,89	-224,92	65,96	0,50	0,99
22	0,0346	-3,36	50,00	11,31	2.500,00	-168,16	66,02	256,57	0,97
23	0,0373	-3,29	78,95	10,81	6.233,10	-259,62	66,41	157,35	0,83
24	0,0381	-3,27	56,52	10,67	3.194,51	-184,66	66,52	99,94	0,79
25	0,0406	-3,20	57,14	10,27	3.264,98	-183,11	66,84	94,11	0,69
26	0,0423	-3,16	81,82	10,01	6.694,51	-258,82	67,06	217,97	0,62
27	0,0452	-3,10	90,00	9,59	8.100,00	-278,76	67,40	510,80	0,52
28	0,0462	-3,07	80,00	9,45	6.400,00	-245,98	67,52	155,84	0,49
29	0,0501	-2,99	50,00	8,96	2.500,00	-149,64	67,94	321,90	0,38
30	0,0542	-2,92	90,00	8,50	8.100,00	-262,36	68,35	468,91	0,29
31	0,0592	-2,83	72,22	7,99	5.215,73	-204,16	68,80	11,67	0,20

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Nr.	VS	$w_i = \ln(VS)$	y_i	w_i^2	y_i^2	$w_i \cdot y_i$	$y_{i0} = y_m + b(w_i - w_m)$	$(y_i - y_{i0})^2$	$(w_i - w_m)^2$
32	0,0610	-2,80	66,67	7,83	4.444,89	-186,50	68,96	5,23	0,18
33	0,0618	-2,78	44,44	7,75	1.974,91	-123,71	69,03	604,60	0,17
34	0,0738	-2,61	68,18	6,80	4.648,51	-177,75	69,95	3,12	0,05
35	0,0751	-2,59	78,57	6,70	6.173,24	-203,39	70,04	72,73	0,04
36	0,0852	-2,46	77,78	6,06	6.049,73	-191,51	70,70	50,15	0,01
37	0,0983	-2,32	70,59	5,38	4.982,95	-163,75	71,44	0,72	0,00
38	0,1083	-2,22	84,62	4,94	7.160,54	-188,11	71,94	160,74	0,02
39	0,1102	-2,21	85,71	4,86	7.346,20	-189,00	72,03	187,01	0,03
40	0,1107	-2,20	82,35	4,84	6.781,52	-181,26	72,06	105,97	0,03
41	0,1182	-2,14	82,35	4,56	6.781,52	-175,87	72,40	99,09	0,06
42	0,1211	-2,11	58,33	4,46	3.402,39	-123,15	72,52	201,41	0,07
43	0,1767	-1,73	78,95	3,00	6.233,10	-136,84	74,49	19,93	0,41
44	0,1924	-1,65	78,57	2,72	6.173,24	-129,50	74,93	13,27	0,53
45	0,2818	-1,27	90,00	1,60	8.100,00	-113,99	76,91	171,33	1,23
46	0,3216	-1,13	90,91	1,29	8.264,63	-103,13	77,60	177,22	1,54
47	0,3434	-1,07	77,78	1,14	6.049,73	-83,14	77,94	0,02	1,71
48	0,4705	-0,75	87,50	0,57	7.656,25	-65,97	79,57	62,82	2,63
49	0,6553	-0,42	77,78	0,18	6.049,73	-32,87	81,30	12,36	3,82
50	0,6912	-0,37	78,57	0,14	6.173,24	-29,02	81,57	9,02	4,03
51	0,7204	-0,33	72,73	0,11	5.289,65	-23,85	81,79	82,04	4,20
52	1,1667	0,15	71,43	0,02	5.102,24	11,01	84,29	165,44	6,41
53	1,2894	0,25	78,57	0,06	6.173,24	19,97	84,81	38,96	6,92
54	1,3210	0,28	76,47	0,08	5.847,66	21,29	84,94	71,70	7,05
55	1,4402	0,36	80,00	0,13	6.400,00	29,18	85,39	29,02	7,52
56	1,4964	0,40	92,86	0,16	8.622,98	37,43	85,59	52,92	7,73
57	1,6009	0,47	84,62	0,22	7.160,54	39,82	85,94	1,73	8,11
58	3,1280	1,14	90,91	1,30	8.264,63	103,67	89,42	2,23	12,37
59	5,6259	1,73	91,67	2,98	8.403,39	158,35	92,47	0,63	16,85
60	5,7881	1,76	83,33	3,08	6.943,89	146,31	92,61	86,19	17,08
61	10,9551	2,39	90,00	5,73	8.100,00	215,44	95,93	35,15	22,76
62	21,0393	3,05	85,71	9,28	7.346,20	261,11	99,32	185,22	29,41
Σ :		<u>-147,37</u>	<u>4.410,77</u>	<u>625,30</u>	<u>331.092,81</u>	<u>-9.055,60</u>		<u>9.880,86</u>	<u>274,99</u>

Für die Summe der Quadrate der Abweichungen der w-Werte von deren Mittelwert gilt:

$$\sum (w_i - w_m)^2 = 274,99 \quad (K.6)$$

Damit werden nach den Formeln (7.15) bzw. (7.11) die Größen C und D für die Bestimmung der Prognose- bzw. Konfidenzintervalle bestimmt. In Spalte (4) bezeichnet t den Wert $t_{60;1-\alpha/2} = 1,671$ das Quantil der t-Verteilung mit $n - 2 = 60$ Freiheitsgraden für das zweiseitige Signifikanzniveau $\alpha = 0,10$. Die Größe s bezeichnet den Schätzwert für die Standardabweichung der Grundgesamtheit. Die Grenzen des Prognoseintervalls $P_{0,oben}$ und $P_{0,unten}$ werden nach den Formeln (7.17) bzw. (7.16) ermittelt. Die Grenzen $C_{0,oben}$ und $C_{0,unten}$ des Konfidenzintervalls werden analog zu diesen Formeln bestimmt.

*Tabelle K.2: Bestimmung der Konfidenz- und Prognoseintervalle
für den Zusammenhang zwischen VS und P*

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Nr.	VS	C	t·s·C	C _{0,oben}	C _{0,unten}	D	t·s·D	P _{0,oben}	P _{0,unten}
1	0,0009	0,31	6,59	53,62	40,44	1,05	22,43	69,46	24,60
2	0,0018	0,27	5,78	56,44	44,88	1,04	22,21	72,87	28,45
3	0,0034	0,24	5,07	59,03	48,90	1,03	22,03	76,00	31,93
4	0,0038	0,23	4,96	59,45	49,53	1,03	22,01	76,50	32,48
5	0,0098	0,19	3,98	63,46	55,50	1,02	21,81	81,29	37,67
6	0,0120	0,18	3,82	64,22	56,59	1,02	21,78	82,18	38,62
7	0,0123	0,18	3,77	64,42	56,87	1,02	21,77	82,42	38,87
8	0,0135	0,17	3,69	64,83	57,45	1,01	21,76	82,90	39,38
9	0,0142	0,17	3,65	65,02	57,72	1,01	21,75	83,13	39,62
10	0,0142	0,17	3,65	65,03	57,73	1,01	21,75	83,13	39,63
11	0,0143	0,17	3,64	65,08	57,80	1,01	21,75	83,19	39,69
12	0,0181	0,16	3,45	66,09	59,19	1,01	21,72	84,36	40,92
13	0,0182	0,16	3,44	66,13	59,24	1,01	21,72	84,40	40,97
14	0,0185	0,16	3,43	66,20	59,34	1,01	21,72	84,48	41,05
15	0,0198	0,16	3,38	66,49	59,74	1,01	21,71	84,82	41,41
16	0,0224	0,15	3,28	67,05	60,48	1,01	21,69	85,46	42,07
17	0,0227	0,15	3,28	67,11	60,56	1,01	21,69	85,52	42,14
18	0,0245	0,15	3,22	67,45	61,00	1,01	21,68	85,91	42,54
19	0,0260	0,15	3,18	67,71	61,35	1,01	21,68	86,21	42,85
20	0,0262	0,15	3,18	67,74	61,38	1,01	21,68	86,24	42,88
21	0,0343	0,14	3,01	68,98	62,95	1,01	21,65	87,62	44,31
22	0,0346	0,14	3,01	69,02	63,01	1,01	21,65	87,67	44,36
23	0,0373	0,14	2,97	69,37	63,44	1,01	21,65	88,05	44,76
24	0,0381	0,14	2,96	69,47	63,56	1,01	21,65	88,16	44,87
25	0,0406	0,14	2,93	69,77	63,92	1,01	21,64	88,48	45,20
26	0,0423	0,14	2,91	69,96	64,15	1,01	21,64	88,70	45,42
27	0,0452	0,13	2,88	70,28	64,52	1,01	21,64	89,03	45,76
28	0,0462	0,13	2,87	70,39	64,65	1,01	21,63	89,15	45,88

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Nr.	VS	C	t·s·C	C _{0,oben}	C _{0,unten}	D	t·s·D	P _{0,oben}	P _{0,unten}
29	0,0501	0,13	2,84	70,78	65,10	1,01	21,63	89,57	46,31
30	0,0542	0,13	2,81	71,16	65,53	1,01	21,63	89,97	46,72
31	0,0592	0,13	2,78	71,59	66,02	1,01	21,62	90,43	47,18
32	0,0610	0,13	2,78	71,73	66,18	1,01	21,62	90,58	47,33
33	0,0618	0,13	2,77	71,80	66,25	1,01	21,62	90,65	47,41
34	0,0738	0,13	2,74	72,69	67,21	1,01	21,62	91,56	48,33
35	0,0751	0,13	2,74	72,78	67,30	1,01	21,62	91,66	48,42
36	0,0852	0,13	2,73	73,42	67,97	1,01	21,62	92,31	49,08
37	0,0983	0,13	2,72	74,16	68,71	1,01	21,62	93,05	49,82
38	0,1083	0,13	2,73	74,67	69,21	1,01	21,62	93,56	50,32
39	0,1102	0,13	2,73	74,77	69,30	1,01	21,62	93,65	50,42
40	0,1107	0,13	2,73	74,79	69,32	1,01	21,62	93,67	50,44
41	0,1182	0,13	2,74	75,14	69,65	1,01	21,62	94,01	50,78
42	0,1211	0,13	2,74	75,27	69,78	1,01	21,62	94,14	50,90
43	0,1767	0,13	2,85	77,33	71,64	1,01	21,63	96,12	52,85
44	0,1924	0,13	2,88	77,81	72,05	1,01	21,64	96,56	53,29
45	0,2818	0,14	3,08	79,99	73,83	1,01	21,66	98,57	55,25
46	0,3216	0,15	3,16	80,76	74,44	1,01	21,68	99,27	55,92
47	0,3434	0,15	3,21	81,14	74,73	1,01	21,68	99,62	56,26
48	0,4705	0,16	3,44	83,01	76,14	1,01	21,72	101,29	57,86
49	0,6553	0,17	3,72	85,01	77,58	1,01	21,76	103,06	59,53
50	0,6912	0,18	3,76	85,34	77,81	1,02	21,77	103,34	59,80
51	0,7204	0,18	3,80	85,59	77,99	1,02	21,78	103,57	60,01
52	1,1667	0,20	4,26	88,55	80,03	1,02	21,86	106,15	62,43
53	1,2894	0,20	4,36	89,17	80,45	1,02	21,88	106,69	62,93
54	1,3210	0,20	4,38	89,32	80,56	1,02	21,89	106,82	63,05
55	1,4402	0,21	4,47	89,86	80,92	1,02	21,90	107,29	63,48
56	1,4964	0,21	4,51	90,10	81,08	1,02	21,91	107,50	63,67
57	1,6009	0,21	4,58	90,52	81,36	1,02	21,93	107,86	64,01
58	3,1280	0,25	5,30	94,72	84,12	1,03	22,09	111,51	67,33
59	5,6259	0,28	5,97	98,43	86,50	1,04	22,26	114,72	70,21
60	5,7881	0,28	6,00	98,61	86,62	1,04	22,27	114,88	70,35
61	10,9551	0,31	6,74	102,67	89,19	1,05	22,48	118,41	73,45
62	21,0393	0,35	7,52	106,84	91,80	1,06	22,73	122,04	76,59

Eine analoge Regressionsrechnung kann auch für den Zusammenhang zwischen den Werten VS und P' durchgeführt werden. Das genaue Vorgehen wird in Abschnitt 7.3 erläutert. Die Ausgangswerte für die Berechnung ergeben sich aus Tabelle K.3.

Tabelle K.3: Zusammenstellung der für die Regressionsrechnung für den Zusammenhang zwischen VS und P' benötigten Ausgangswerte

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Nr.	VS	$w_i = \ln(VS)$	y_i	w_i^2	y_i^2	$w_i \cdot y_i$	$y_{i0} = y_m + b(w_i - w_m)$	$(y_i - y_{i0})^2$	$(w_i - w_m)^2$
1	0,0009	-7,02	46,15	49,25	2129,82	-323,88	63,58	303,91	21,54
2	0,0018	-6,32	42,86	39,94	1836,98	-270,86	66,53	560,02	15,55
3	0,0034	-5,68	75,00	32,30	5625,00	-426,23	69,21	33,57	10,93
4	0,0038	-5,58	64,29	31,16	4133,20	-358,87	69,63	28,54	10,27
5	0,0098	-4,62	77,27	21,36	5970,65	-357,16	73,68	12,92	5,04
6	0,0120	-4,44	47,06	19,75	2214,64	-209,13	74,43	748,91	4,27
7	0,0123	-4,40	80,00	19,34	6400,00	-351,78	74,62	28,91	4,08
8	0,0135	-4,30	70,59	18,50	4982,95	-303,64	75,03	19,68	3,70
9	0,0142	-4,26	66,67	18,12	4444,89	-283,82	75,21	72,99	3,54
10	0,0142	-4,26	71,43	18,11	5102,24	-303,95	75,22	14,37	3,53
11	0,0143	-4,25	80,00	18,02	6400,00	-339,60	75,26	22,43	3,49
12	0,0181	-4,01	68,42	16,10	4681,30	-274,58	76,24	61,17	2,68
13	0,0182	-4,00	68,42	16,04	4681,30	-274,01	76,28	61,71	2,65
14	0,0185	-3,99	93,33	15,91	8710,49	-372,27	76,34	288,53	2,60
15	0,0198	-3,92	100,00	15,38	10000,00	-392,15	76,63	546,29	2,39
16	0,0224	-3,80	80,00	14,42	6400,00	-303,79	77,15	8,12	2,02
17	0,0227	-3,78	59,09	14,32	3491,63	-223,59	77,21	328,21	1,98
18	0,0245	-3,71	100,00	13,75	10000,00	-370,85	77,52	505,16	1,77
19	0,0260	-3,65	85,00	13,32	7225,00	-310,16	77,78	52,20	1,62
20	0,0262	-3,64	80,00	13,27	6400,00	-291,47	77,80	4,85	1,60
21	0,0343	-3,37	80,00	11,38	6400,00	-269,89	78,94	1,13	0,99
22	0,0346	-3,36	71,43	11,31	5102,24	-240,23	78,98	56,99	0,97
23	0,0373	-3,29	84,21	10,81	7091,32	-276,92	79,29	24,17	0,83
24	0,0381	-3,27	56,52	10,67	3194,51	-184,66	79,38	522,75	0,79
25	0,0406	-3,20	64,29	10,27	4133,20	-206,03	79,65	235,83	0,69
26	0,0423	-3,16	90,91	10,01	8264,63	-287,57	79,82	122,96	0,62
27	0,0452	-3,10	100,00	9,59	10000,00	-309,73	80,10	396,05	0,52
28	0,0462	-3,07	80,00	9,45	6400,00	-245,98	80,19	0,04	0,49
29	0,0501	-2,99	58,33	8,96	3402,39	-174,58	80,54	493,23	0,38
30	0,0542	-2,92	90,00	8,50	8100,00	-262,36	80,87	83,42	0,29
31	0,0592	-2,83	88,89	7,99	7901,43	-251,29	81,24	58,56	0,20
32	0,0610	-2,80	66,67	7,83	4444,89	-186,50	81,36	215,87	0,18
33	0,0618	-2,78	100,00	7,75	10000,00	-278,37	81,42	345,21	0,17
34	0,0738	-2,61	77,27	6,80	5970,65	-201,44	82,16	23,95	0,05
35	0,0751	-2,59	85,71	6,70	7346,20	-221,88	82,24	12,03	0,05
36	0,0852	-2,46	88,89	6,06	7901,43	-218,87	82,77	37,40	0,01
37	0,0983	-2,32	82,35	5,38	6781,52	-191,04	83,37	1,05	0,00

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Nr.	VS	$w_i = \ln(VS)$	y_i	w_i^2	y_i^2	$w_i \cdot y_i$	$y_{i0} = y_m + b(w_i - w_m)$	$(y_i - y_{i0})^2$	$(w_i - w_m)^2$
38	0,1083	-2,22	84,62	4,94	7160,54	-188,11	83,78	0,70	0,02
39	0,1102	-2,21	100,00	4,86	10000,00	-220,51	83,86	260,58	0,03
40	0,1107	-2,20	88,24	4,84	7786,30	-194,22	83,87	19,06	0,03
41	0,1182	-2,14	82,35	4,56	6781,52	-175,87	84,15	3,24	0,06
42	0,1211	-2,11	100,00	4,46	10000,00	-211,13	84,25	247,98	0,07
43	0,1767	-1,73	94,74	3,00	8975,67	-164,21	85,85	79,12	0,41
44	0,1924	-1,65	100,00	2,72	10000,00	-164,82	86,20	190,35	0,53
45	0,2818	-1,27	95,00	1,60	9025,00	-120,33	87,81	51,68	1,23
46	0,3216	-1,13	100,00	1,29	10000,00	-113,44	88,37	135,31	1,54
47	0,3434	-1,07	100,00	1,14	10000,00	-106,89	88,64	128,97	1,71
48	0,4705	-0,75	100,00	0,57	10000,00	-75,39	89,97	100,60	2,63
49	0,6553	-0,42	88,89	0,18	7901,43	-37,57	91,37	6,13	3,82
50	0,6912	-0,37	92,86	0,14	8622,98	-34,30	91,59	1,61	4,03
51	0,7204	-0,33	90,91	0,11	8264,63	-29,81	91,77	0,73	4,20
52	1,1667	0,15	85,71	0,02	7346,20	13,21	93,80	65,38	6,41
53	1,2894	0,25	100,00	0,06	10000,00	25,42	94,22	33,44	6,92
54	1,3210	0,28	94,12	0,08	8858,57	26,20	94,32	0,04	7,05
55	1,4402	0,36	100,00	0,13	10000,00	36,48	94,68	28,27	7,52
56	1,4964	0,40	92,86	0,16	8622,98	37,43	94,84	3,94	7,73
57	1,6009	0,47	84,62	0,22	7160,54	39,82	95,13	110,43	8,11
58	3,1280	1,14	90,91	1,30	8264,63	103,67	97,95	49,56	12,37
59	5,6259	1,73	100,00	2,98	10000,00	172,74	100,42	0,18	16,85
60	5,7881	1,76	91,67	3,08	8403,39	160,95	100,54	78,72	17,08
61	10,9551	2,39	90,00	5,73	8100,00	215,44	103,23	175,04	22,76
62	21,0393	3,05	85,71	9,28	7346,20	261,11	105,98	410,84	29,41
Σ :		<u>-147,37</u>	<u>5.154,26</u>	<u>625,30</u>	<u>441.885,12</u>	<u>-11.093,22</u>		<u>8.515,05</u>	<u>274,99</u>

Es ergeben sich unter Anwendung der o.a. Formeln aus den hier tabellierten Werten:

$$r = 0,604$$

$$b = 4,21$$

$$a = 93,15$$

$$s = 11,91$$

Unter Berücksichtigung des Quantils $t_{60;1-\alpha/2} = 1,671$ als t ergeben sich damit die in Tabelle K.4 enthaltenen Grenzwerte für die Konfidenz- und Prognoseintervalle.

*Tabelle K.4: Bestimmung der Konfidenz- und Prognoseintervalle
für den Zusammenhang zwischen VS und P'*

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Nr.	VS	C	t·s·C	C _{0,oben}	C _{0,unten}	D	t·s·D	P _{0,oben}	P _{0,unten}
1	0,0009	0,31	6,12	69,70	57,47	1,05	20,83	84,41	42,76
2	0,0018	0,27	5,37	71,89	61,16	1,04	20,62	87,14	45,91
3	0,0034	0,24	4,71	73,91	64,50	1,03	20,46	89,66	48,75
4	0,0038	0,23	4,60	74,24	65,03	1,03	20,43	90,06	49,20
5	0,0098	0,19	3,70	77,37	69,98	1,02	20,25	93,92	53,43
6	0,0120	0,18	3,54	77,97	70,88	1,02	20,22	94,65	54,21
7	0,0123	0,18	3,50	78,13	71,12	1,02	20,21	94,84	54,41
8	0,0135	0,17	3,43	78,45	71,60	1,02	20,20	95,23	54,83
9	0,0142	0,17	3,39	78,60	71,83	1,01	20,19	95,41	55,02
10	0,0142	0,17	3,39	78,61	71,83	1,01	20,19	95,41	55,03
11	0,0143	0,17	3,38	78,64	71,89	1,01	20,19	95,46	55,07
12	0,0181	0,16	3,20	79,44	73,04	1,01	20,16	96,40	56,08
13	0,0182	0,16	3,20	79,47	73,08	1,01	20,16	96,44	56,11
14	0,0185	0,16	3,18	79,53	73,16	1,01	20,16	96,50	56,18
15	0,0198	0,16	3,14	79,76	73,49	1,01	20,15	96,78	56,48
16	0,0224	0,15	3,05	80,20	74,10	1,01	20,14	97,29	57,01
17	0,0227	0,15	3,04	80,25	74,17	1,01	20,14	97,34	57,07
18	0,0245	0,15	2,99	80,52	74,53	1,01	20,13	97,65	57,39
19	0,0260	0,15	2,95	80,73	74,82	1,01	20,12	97,90	57,65
20	0,0262	0,15	2,95	80,75	74,85	1,01	20,12	97,92	57,68
21	0,0343	0,14	2,80	81,73	76,14	1,01	20,10	99,04	58,83
22	0,0346	0,14	2,79	81,77	76,19	1,01	20,10	99,08	58,88
23	0,0373	0,14	2,76	82,05	76,54	1,01	20,10	99,39	59,20
24	0,0381	0,14	2,75	82,13	76,64	1,01	20,10	99,48	59,29
25	0,0406	0,14	2,72	82,36	76,93	1,01	20,09	99,74	59,56
26	0,0423	0,14	2,70	82,52	77,12	1,01	20,09	99,91	59,73
27	0,0452	0,13	2,67	82,77	77,43	1,01	20,09	100,18	60,01
28	0,0462	0,13	2,66	82,86	77,53	1,01	20,08	100,28	60,11
29	0,0501	0,13	2,63	83,17	77,91	1,01	20,08	100,62	60,46
30	0,0542	0,13	2,61	83,48	78,26	1,01	20,08	100,94	60,79
31	0,0592	0,13	2,59	83,82	78,65	1,01	20,07	101,31	61,16
32	0,0610	0,13	2,58	83,94	78,79	1,01	20,07	101,44	61,29
33	0,0618	0,13	2,58	84,00	78,85	1,01	20,07	101,49	61,35
34	0,0738	0,13	2,54	84,71	79,62	1,01	20,07	102,23	62,10
35	0,0751	0,13	2,54	84,78	79,70	1,01	20,07	102,31	62,17
36	0,0852	0,13	2,53	85,30	80,24	1,01	20,07	102,84	62,71
37	0,0983	0,13	2,53	85,90	80,85	1,01	20,07	103,44	63,31

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Nr.	VS	C	t·s·C	C _{0,oben}	C _{0,unten}	D	t·s·D	P _{0,oben}	P _{0,unten}
38	0,1083	0,13	2,54	86,32	81,25	1,01	20,07	103,85	63,72
39	0,1102	0,13	2,54	86,39	81,32	1,01	20,07	103,93	63,79
40	0,1107	0,13	2,54	86,41	81,34	1,01	20,07	103,94	63,81
41	0,1182	0,13	2,55	86,70	81,61	1,01	20,07	104,22	64,08
42	0,1211	0,13	2,55	86,80	81,70	1,01	20,07	104,32	64,18
43	0,1767	0,13	2,64	88,49	83,20	1,01	20,08	105,93	65,76
44	0,1924	0,13	2,68	88,88	83,53	1,01	20,09	106,29	66,12
45	0,2818	0,14	2,86	90,67	84,95	1,01	20,11	107,92	67,70
46	0,3216	0,15	2,94	91,30	85,43	1,01	20,12	108,49	68,25
47	0,3434	0,15	2,98	91,62	85,67	1,01	20,13	108,77	68,52
48	0,4705	0,16	3,19	93,16	86,78	1,01	20,16	110,13	69,81
49	0,6553	0,17	3,45	94,82	87,92	1,02	20,20	111,57	71,16
50	0,6912	0,18	3,49	95,08	88,10	1,02	20,21	111,80	71,38
51	0,7204	0,18	3,53	95,29	88,24	1,02	20,22	111,98	71,55
52	1,1667	0,20	3,95	97,75	89,84	1,02	20,30	114,09	73,50
53	1,2894	0,20	4,05	98,26	90,17	1,02	20,31	114,53	73,90
54	1,3210	0,20	4,07	98,39	90,25	1,02	20,32	114,64	74,00
55	1,4402	0,21	4,15	98,83	90,53	1,02	20,33	115,02	74,35
56	1,4964	0,21	4,19	99,03	90,66	1,02	20,34	115,19	74,50
57	1,6009	0,21	4,25	99,38	90,88	1,02	20,36	115,48	74,77
58	3,1280	0,25	4,92	102,87	93,03	1,03	20,51	118,46	77,44
59	5,6259	0,28	5,54	105,96	94,89	1,04	20,66	121,09	79,76
60	5,7881	0,28	5,57	106,11	94,98	1,04	20,67	121,21	79,87
61	10,9551	0,31	6,26	109,49	96,97	1,05	20,87	124,10	82,36
62	21,0393	0,35	6,98	112,96	99,00	1,06	21,10	127,08	84,88

Lebenslauf

Name	Nils Schnittker
geboren	1975 in Bremen
1994 – 2000	Studium an der TU Hamburg-Harburg im Studiengang „Bauingenieurwesen und Umwelttechnik“, Abschluss als Diplom-Ingenieur
2001-2006	wissenschaftlicher Mitarbeiter am Institut für Massivbau der TU Darmstadt, Fachgebiet Baustoffe, Bauphysik, Bauchemie
seit 2006	selbstständig als Partner der „Gesellschaft für Weiterbildung im Bauwesen GbR“ in Darmstadt